A Survey on Wireless Device-free Human Sensing: Application Scenarios, Current Solutions, and Open Issues

JIANG XIAO, HUICHUWU LI, MINRUI WU, and HAI JIN, National Engineering Research Center for Big Data Technology and System, Services Computing Technology and System Lab, Cluster and Grid Computing Lab, School of Computer Science and Technology, Huazhong University of Science and Technology, China
M. JAMAL DEEN, McMaster University, Canada
JIANNONG CAO, The Hong Kong Polytechnic University, China

In the last decade, many studies have significantly pushed the limits of *wireless device-free human sensing* (WDHS) technology and facilitated various applications, ranging from activity identification to vital sign monitoring. This survey presents a novel taxonomy that classifies the state-of-the-art WDHS systems into 11 categories according to their *sensing task type* and motion *granularity*. In particular, existing WDHS systems involve three primary sensing task types. The first type, *behavior recognition*, is a classification problem of recognizing predefined meaningful behaviors. The second type is *movement tracking*, monitoring the quantitative values of behavior states integrating with spatiotemporal information. The third type, *user identification*, leverages the unique features in behaviors to identify who performs the movements. The selected papers in each sensing task type can be further divided into sub-categories according to their motion granularity. Recent advances reveal that WDHS systems within a particular granularity follow similar challenges and design principles. For example, fine-grained hand recognition systems target extracting subtle motion-induced signal changes from the noisy signal responses, and their sensing areas are limited to a relatively small range. Coarse-grained activity identification systems need to overcome the interference of other moving objects within the room-level sensing range. A novel research framework is proposed to help to summarize WDHS systems from methodology, evaluation performance, and design goals. Finally, we conclude with several open issues and present the future research directions from the perspectives of *data collection*, *sensing methodology*, *performance evaluation*, and *application scenario*.

CCS Concepts: • **General and reference → Surveys and overviews**; • **Human-centered computing → Ubiquitous and mobile computing systems and tools**; **Ubiquitous and mobile computing design and evaluation methods;**

Additional Key Words and Phrases: Human sensing, wireless, device-free, application scenario, sensing task type, granularity

This work was supported in part by the Technology Innovation Project of Hubei Province of China under grant No. 2019AEA171, and Key Research and Development Program of Hubei Province No. 2021BEA164.
Authors' addresses: J. Xiao, H. Li, M. Wu, and H. Jin, National Engineering Research Center for Big Data Technology and System, Services Computing Technology and System Lab, Cluster and Grid Computing Lab, School of Computer Science and Technology, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan, Hubei, China, 430074; emails: {jiangxiao, credolee, hustwmr, hjin}@hust.edu.cn; M. J. Deen, McMaster University, Hamilton, Ontario, Canada; J. Cao, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China.

0360-0300/2022/12-ART88 $15.00
https://doi.org/10.1145/3530682

ACM Reference format:
Jiang Xiao, Huichuwu Li, Minrui Wu, Hai Jin, M. Jamal Deen, and Jiannong Cao. 2022. A Survey on Wireless Device-free Human Sensing: Application Scenarios, Current Solutions, and Open Issues. *ACM Comput. Surv.* 55, 5, Article 88 (December 2022), 35 pages.
https://doi.org/10.1145/3530682

1 INTRODUCTION

With the emergence of the *Internet of Things*, human sensing techniques have attracted considerable attention in creating a smart life with many compelling applications, e.g., smart healthcare, safety surveillance, and ubiquitous interaction. To promote these applications, researchers have explored many sensing techniques [10, 35, 53] to obtain spatiotemporal and motion-related information of human behaviors. For example, in a construction site scene, a sensing system should have the ability to track workers to alert them when they enter dangerous areas. When an elderly person living alone falls, a sensing system with a fall detection function can promptly notify their children or guardians for rescue.

Human sensing techniques can be classified into two categories: device-based and device-free. The device-based approach depends on wearable sensors [49, 55] to monitor human health and behaviors and achieve seamless availability across different environments. It is still unsuitable for some scenarios, e.g., elderly patient monitoring. On-body sensors cause additional carrying burdens, and people may forget to wear them. Device-free human sensing, namely ***wireless device-free human sensing*** **(WDHS)**, provides a non-intrusive sensing approach. WDHS utilizes the infrastructure, e.g., camera [112], signal readers [81], access points [86], and radar [119], to capture human motions.

The image-based approaches have achieved a remarkable breakthrough in the last decade and enabled many mature commercial applications. However, they are limited by some inherent problems. First, most of the systems require that the target persons not be impeded. Second, they are sensitive to the brightness of the environment. Third, cameras make people feel monitored continuously by some unknown people. These problems make the image-based techniques unsuitable for some application scenarios focusing on long-term and privacy-conscious sensing tasks.

Researchers have proposed many attractive solutions involving various wireless signals such as Wi-Fi [61], ***Radio Frequency Identification*** **(RFID)** [29], ultrasonic [68], and ***visible light*** **(VL)** [111] to broaden the boundaries of WDHS. These heterogeneous wireless signals enable fantastic applications in a variety of scenarios such as through-wall sensing [100], whole-home intrusion detection [113], and vital sign monitoring [83]. Since there are many studies based on different techniques, a scoping review is required to help researchers efficiently understand the research state.

To date, several surveys relevant to WDHS have been published, as summarized in Table 1. Systematic surveys provide an in-depth analysis of particular narrow scopes, e.g., application scenarios [16, 34, 70], wireless techniques [46], and sensing methods [35, 58], since they can comprehensively compare the relevant systems.

Unlike the systematic surveys, the scoping surveys provide a broader scope of WDHS. For example, Ngamakeur et al. [56] reviewed the techniques of device-free indoor localization and tracking. Liu et al. [42] surveyed the wireless sensing techniques for multiple application scenarios. Compared with these two scoping reviews, this survey reviews the WDHS systems involving more wireless techniques and application scenarios.

We conducted an initial search of scientific databases, e.g., Google Scholar, ACM Digital Library, Web of Science, and IEEE Xplore, to find the most related literature using combination keywords:

Table 1. Summary of Related Surveys on WDHS

Reference	Type	Application Scenarios	Signal Sources	Topic & Taxonomy
Shit et al. [70]	Systematic	Whole-body movement tracking	RFID, FM, ultrasonic, and Wi-Fi	Techniques of device-free localization
Kouyoumdjieva et al. [34]	Systematic	Whole-body movement tracking	Acoustic, Wi-Fi, and VL	Techniques of crowd counting
Deep et al. [16]	Systematic	Activity identification	RFID and Wi-Fi	Techniques of anomalous behavior detection for elderly care
Ma et al. [46]	Systematic	Activity identification, hand gesture recognition, whole-body movement tracking, hand/finger tracking, vital sign monitoring, and whole-body-motion-based authentication	Wi-Fi	Techniques of CSI-based human sensing
Nirmal et al. [58]	Systematic	Activity identification, hand gesture recognition, whole-body movement tracking, pose estimation, vital sign monitoring, and whole-body-motion-based authentication	RFID, FM, and Wi-Fi	Deep learning techniques of device-free human sensing
Li et al. [35]	Systematic	Activity identification, hand gesture recognition, whole-body movement tracking, pose estimation, hand/finger tracking, vital sign monitoring, and whole-body-motion-based authentication	RFID, FM, and Wi-Fi	Deep learning techniques for wireless sensing
Ngamakeur et al. [56]	Scoping	Activity identification, whole-body movement tracking, whole-body-motion-based authentication	Acoustic, RFID, and Wi-Fi	Techniques of device-free indoor localization for multi-resident scenarios
Liu et al. [42]	Scoping	Activity identification, hand gesture recognition, whole-body movement tracking, vital sign monitoring, and whole-body-motion-based authentication	FM and Wi-Fi	Techniques of wireless sensing
This survey	Scoping	Activity identification, limb motion recognition, hand gesture recognition, lip reading, whole-body movement tracking, pose estimation, hand/finger tracking, vital sign monitoring, whole-body-motion-based authentication, finger-motion-based authentication, and lip-motion-based authentication	Audio, ambient *radio frequency* (RF), EMF, FM, RFID, mmWave, solar, ultrasonic, VL, Wi-Fi	A scoping review of the application of WDHS according to the sensing task type and motion granularity

wireless, device-free, human sensing, and gesture recognition. Then, we inspected the proceedings of the major relevant conferences (e.g., AAAI, CCS, CHI, CVPR, ICDCS, INFOCOM, IPSN, MobiCom, MobiHoc, MobiSys, NSDI, PerCom, SECON, SenSys, SIGCOMM, UbiComp, and WWW) and the papers of the related transactions (e.g., IMWUT and TMC) from 2013. Next, a novel taxonomy is proposed to classify the recent advances of WDHS systems according to their sensing task types and granularity.

1.1 Taxonomy of WDHS

We propose a novel taxonomy that classifies WDHS systems according to *sensing task type* and motion *granularity*, as shown in Figure 1. After reviewing the selected papers, current WDHS systems can be classified into three classes according to their sensing task types: *behavior recognition*, *movement tracking*, and *user identification*. Such classes can be further classified into sub-categories according to their granularity. Previous research [66] showed that motion granularity affects sensing accuracy and brings different challenges. For example, WDHS systems with small sensing targets such as fingers [57] and lips [44] are often limited to a small sensing range to capture slight movements. The short sensing range constrains the application scenario to a single user and requires a specific deployment setting. In contrast, WDHS systems designed for larger sensing targets, such as whole-body movements, have a broader sensing range. However, they require more attention to mitigate the unavoidable interference caused by other people in practical applications.

Behavior recognition is a pure classification problem of recognizing predefined meaningful gestures or activities [52]. It has two main challenges, which are separating target behaviors from ambient noise and designing a proper classification model to distinguish the differences among the detected behaviors. When people move in the field covered by wireless signals, their movements will change the propagation paths causing specific signal patterns received by the antennas. The

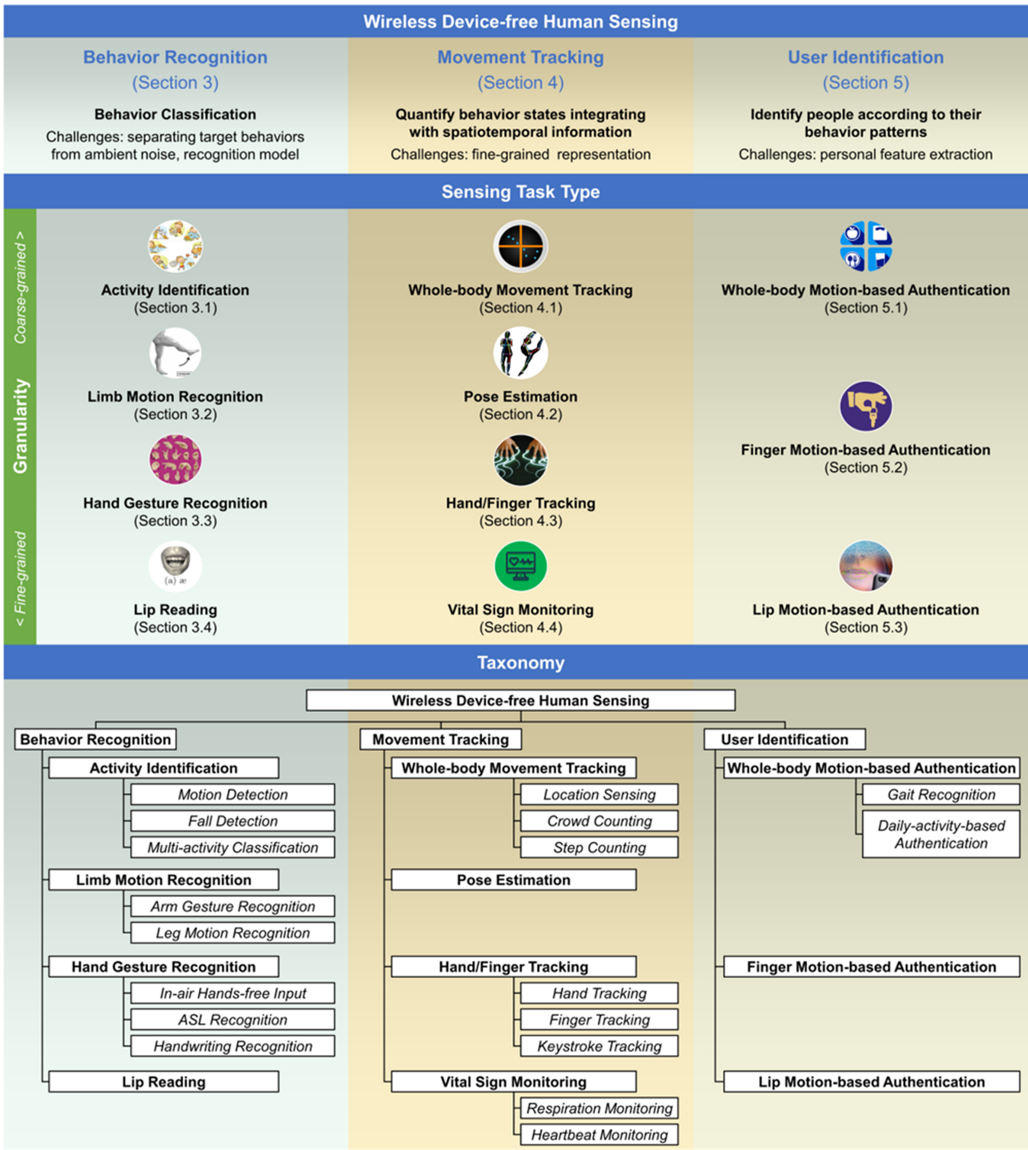


Fig. 1. Taxonomy of wireless device-free human sensing systems.

sensing systems utilize thresholds or pattern fitting methods to detect movement-related segments from the signal streams. Then, these systems extract the features of the segments and input them to machine learning models and deep neural networks for recognition. Following this general framework, many studies were proposed for various sensing targets, including the whole body [61], limbs [79], hands [68], fingers [75], and lips [74], facilitating a wide range of applications.

Movement tracking systems monitor the quantitative values of behavior states integrated with spatiotemporal information. The most challenging problem for these systems is to find a proper representation reflecting the details of the target behaviors. According to the target size, these systems can be classified into whole-body movement tracking [2], pose estimation [117], hand/finger tracking [9], and vital sign monitoring [83]. Typically, location sensing systems [2, 3] aim at estimating the coordinates of the target users who are regarded as points in the maps. Pose estimation is defined as locating human joints [119] like the concept used in the computer vision community [77]. Compared with image-based pose estimation systems, wireless-based methods have excellent performance in through-wall conditions. Hand/finger tracking systems track the hands' locations near smart devices in terms of breaking the limits of small screens. Researchers have explored the use of wireless-based vital sign monitoring [4], which frees users from the on-body sensors.

User identification is to determine who performs the movements. Each person's movement patterns are unique due to the difference in body shapes, behaviors, and health conditions. Such personal information also exists in the wireless signal patterns reflected by the human body, e.g., gaits [85, 109], fingers [31], and lips [44, 74]. However, it is non-trivial to define or extract personal information through a theoretical model. Therefore, existing works leverage experience-based models for user identification.

1.2 Contributions

- We conduct a thorough scoping review of WDHS according to the sensing target types and motion granularity, hoping to help readers keep up-to-date about the research state and inspire more exciting application scenarios.
- We propose a novel researcher framework to guide the summary of WDHS systems from seven components: data collection, preprocessing, segmentation, representation, modeling, evaluation, and application.
- We conclude with the problems of WDHS and present the future research directions from the aspects of *data collection*, *sensing methodology*, *performance evaluation*, and *application scenario*.

The rest of this article is as follows. In Section 2, we introduce the general research framework of WDHS. In Sections 3 to 5, we describe behavior recognition, movement tracking, and user identification. Then, in Section 6, we present the challenges in this area and give an outlook of the research directions. Finally, we provide conclusions in Section 7.

2 RESEARCH FRAMEWORK

In this section, we give a research framework of human sensing from a high-level perspective like the prior work [107]. Since the research framework proposed by the prior work is very general, it is unable to guide us in building up summarization tables of the selected WDHS papers. Therefore, we propose a novel research framework to provide an overview of how the systems process wireless signals for human sensing and how the researchers evaluate their works, as shown in Figure 2. Please note that this article does not describe the detailed methodologies, such as how a system transforms wireless signals from the time domain to the frequency domain. For more details about the sensing methodology of WDHS systems, we recommend the readers refer to the prior surveys mentioned in Table 1 or to the referenced papers. The general research framework consists of seven sequential components: data collection, preprocessing, segmentation, representation, modeling, evaluation, and application.

2.1 Data Collection

The selected WDHS systems involve various wireless techniques, including acoustic signals [96], ***electromagnetic field* (EMF)** [117], ***frequency-modulated* (FM)** radio [3], ***microwave* (mmWave)** [69], RFID [29], VL [102], and Wi-Fi [113].

According to the way of capturing these signals, the selected papers can be divided into two types: *active* and *passive*. The active ones borrow the idea of radar that they have specific patterns

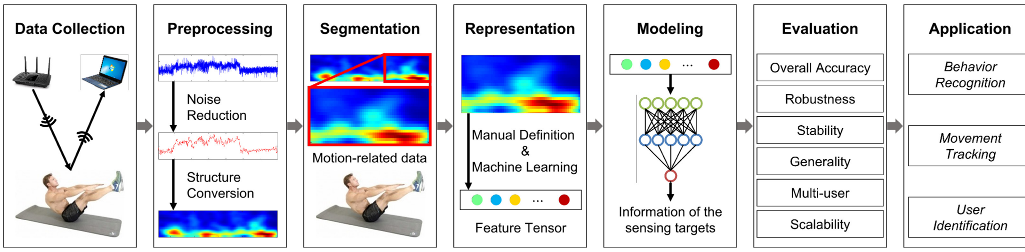


Fig. 2. Research framework of wireless device-free human sensing.

of sensing signals and close distance between transmitting and receiving antennas. Passive sensing systems do not change the device deployment design for communication. They analyze the motion-induced multipath effects to recognize human behaviors.

Most systems extract features from the measurements relevant to signal strength, e.g., ***received signal strength*** **(RSS)** [69] and ***received signal strength indicator*** **(RSSI)** [1]. Such measurements are the superposition of signals of multiple propagation paths. It results in an unstable performance in a dynamic environment and has a coarse-grained spatial resolution. Therefore, researchers introduce phase [90] and ***channel state information*** **(CSI)** [89] that can describe the state of propagation channels in detail.

2.2 Preprocessing

Preprocessing has two major goals of *noise reduction* and *structure conversion*. The meaningful signals are often buried in the noise floor. The noise can be briefly summarized into three types: device-induced noise, environmental noise, and dynamic noise.

To mitigate the device-induced noise, researchers induce an efficient method, namely conjugated multiplication [37]. The key insight is that receiving antennas sharing the same processing circuits have identical measurement errors. The environmental noise refers to the signals with strong strengths reflected by the static environment, having a very low ***Doppler frequency shift*** **(DFS)**. Researchers remove such noise using a ***band-pass filter*** **(BPF)** [65], a ***high-pass filter*** **(HPF)** [69], or background subtraction methods [86]. The dynamic noise is caused by moving objects around the target users. Some moving objects have different moving speeds causing out-band noise, which can be removed by a ***low-pass filter*** **(LPF)**, BPF, HPF, or the combination of ***discrete wavelet transform*** **(DWT)** and thresholding-based methods [75] in the frequency domain. Other moving objects may have a speed close to the target users but are in a farther place. Designers can leverage *long delay removal* [92] to filter out this noise in the time domain.

Besides the mentioned three noise types, other occasional measurement errors, e.g., missing and burst values, degrade the sensing performance. Researchers have explored many methods such as *moving average* [115], *Savitzky-Golay filter* [116], *Hampel filter* [27], ***principal component analysis*** **(PCA)** [86], and *interpolation* [84] to smooth the received signals.

After removing the noise, some systems directly extract features from the data in the time domain. Other systems will convert the data into the frequency domain like spectrogram through ***short-time Fourier transform*** **(STFT)** [79] or DWT [1]. Compared with the raw wireless signals received, the spectrogram provides information on motion speed.

2.3 Segmentation

The goal of this component is to capture candidate signal segments with respect to human behaviors. Many systems use *thresholding* [17] methods to separate motion-induced signals from the background. Some systems leverage *preamble* gestures [80] to notify the sensing systems that the following activities are the targets. Both thresholding-based and preamble-based segmentation methods have inherent delays since they detect motions that are already finished. For real-time scenarios such as movement tracking applications, most systems adopt a *time window* [100] to divide the signal stream into sequences of frames.

2.4 Representation

After segmentation, researchers extract features from the segments for dimension reduction and format alignment. There are three ways to obtain a representation: *pattern encoding*, *feature extraction*, and ***deep neural network* (DNN)**. Pattern encoding uses a series of numbers to encode the instance according to trends in the signal change [1]. In feature extraction, the instances are transformed into manually defined feature vectors, e.g., statistical features [94], ***Mel-Frequency Cepstrum Coefficients* (MFCCs)** [106], profile feature [22, 89], ***Time-of-flight* (ToF)** [9], ***angle-of-arrival* (AoA)** [72], and ***time difference of arrival* (TDoA)** [122]. For complex gestures or activities, it is challenging to determine the features manually. Therefore, DNN models were applied for feature extraction thanks to their strong representation ability [27, 99].

2.5 Modeling

Behavior recognition and user identification systems utilize classification models to distinguish their sensing targets. Based on the complexity of the sensing targets, the sensing models can be divided into two categoires: logic-based model and experience-based model. Logic-based models include *pattern matching* [61] and *if-then-else* statements [29]. These models are computation efficient but only available for tasks with simple movements and small target spaces. For complex sensing targets, like ***American sign language* (ASL)** [69], the corresponding systems prefer to utilize experience-based models, e.g., ***hidden Markov model* (HMM)** [86], ***support vector machine* (SVM)** [22], ***k-nearest neighbor* (kNN)** [80], and DNN [25, 121]. The performance of such experience-based models relies on their training datasets. They become computationally time intensive when the behavior types increase.

Most movement tracking systems concentrate on the coordinates of the sensing targets. Such sensing tasks can design geometric-based mathematical mapping functions [64] to estimate the coordinates or introduce the traditional localization models, such as *dead reckoning* [63], *ellipse model* [3], *triangulation model* [105], and *hyperbola model* [122]. Researchers propose to utilize deep learning models [119, 120] for the pose estimation systems that tackle more complex problems of locating multiple joints.

Other movement tracking systems are designed for specific movements or behaviors, e.g., crowd counting, step counting, and vital sign monitoring. Researchers can more easily find the correlations between such sensing targets and the signal patterns. For example, there is a monotonic relationship between CSI variation and the number of moving people, and human gaits, respiration, and heartbeats have repeated patterns.

2.6 Evaluation

Generally speaking, researchers evaluate their work from the micro and macro perspectives. In the micro view, researchers conduct experiments to find the best parameters or settings based on the details of the method, which we do not summarize in this survey. In the macro perspective, the experiments are conducted to evaluate the practicability of the WDHS systems. In this survey, we summarize the performance evaluation from six aspects, i.e., *overall accuracy*, *robustness*, *stability*, *generality*, *multiuser*, and *scalability*.

- The *overall accuracy* describes the performance for the best experimental settings. We use the term "overall" to denote this performance in the summarization tables. There are two formats for classification and tracking tasks. For classification tasks, including behavior recognition and user identification, researchers use the correct recognition ratio to record the possibility of a WDHS system that can accurately recognize human behaviors or user identity. The movement tracking systems often utilize estimation error to describe the distance between the outputs and the ground truth.
- *Robustness* evaluates the performance when applying a WDHS system to different multipath conditions. The relevant experiments involve the accuracy in different environments, comparison between ***line of sight*** **(LOS)** and ***non-line of sight*** **(NLOS)**, and user location changes.
- *Stability* indicates the ability to overcome continuous signal changes caused by dynamic environments due to unknown persons' activities. In practice, unknown persons occasionally appear within the sensing range when a WDHS system interacts with a registered user. The signals reflected from such unknown persons will confuse the WDHS system. In addition, human activities gradually change the layout. After a long period, the multipath state may be significantly different from the raw condition. Therefore, a practical WDHS system should keep stable performance in long-time-duration applications.
- *Generality* evaluates the sensing performance across different users without additional training costs. Due to the difference in body shapes, behaviors, and health conditions, namely user diversity, a well-trained WDHS system may have poor performance for new users. It requires new users to provide training data for fine-tuning, which seriously influences user experience.
- *Multiuser* records the maximum number of coexisting users. It is an imperative feature for real-world deployment. On the one hand, it is impossible to avoid situations where multiple targets exist at the same time in practical applications, such as room-level location sensing. On the other hand, multi-user support needs to combine the user's location and identity information to make the perception system smarter and safer.
- *Scalability* indicates the size of the sensing target space, reflecting the sensing boundaries of WDHS systems. This survey records the target number of classification systems and the sensing range of tracking systems.

2.7 Application

WDHS systems are used in a wide range of applications. This survey concludes three main categories based on the sensing task types: behavior recognition, movement tracking, and user identification. Behavior recognition and movement tracking are motion-centric sensing tasks, and the third type focuses on user identity. These application scenarios will be discussed in detail in the next sections.

3 BEHAVIOR RECOGNITION

In this section, we present the behavior recognition systems according to the behavior granularity. The selected systems can be classified into four categories: *activity identification*, *limb motion recognition*, *hand gesture recognition*, and *lip reading*. Then, we conclude with the challenges and research opportunities in terms of research guidance at the end of this section.

3.1 Activity Identification

In the past decade, activity identification was one of the hottest research topics in WDHS. It focuses on whole-body movements, promoting exciting applications, such as *motion detection*, *fall*

Table 2. Methodology of Activity Identification Systems

Reference	Data Collection			Preprocessing	Segmentation	Representation	Sensing Model
	Source	Measurement	Type				
[113]	Wi-Fi	Amplitude of CSI	Passive	-	Time Window	Auto-correlation	Hypothesis Testing
[37]	Wi-Fi	CSI	Passive	Conjugate Multiplication	Time Window	Statistical Feature	Thresholding
[22]	Wi-Fi	Amplitude of CSI	Passive	Moving Average	Local Outlier Factor	Statistical Feature, Profile Feature	Random Forests, SVM
[84]	Wi-Fi	CSI	Passive	Interpolation, BPF	Thresholding, Time Window	Statistical Feature, Profile Feature	v-SVM
[115]	Wi-Fi	Amplitude of CSI	Passive	Moving Average, BPF, PCA	Thresholding	Statistical Feature, Profile Feature	Transfer Learning, SVM
[89]	Wi-Fi	CSI	Passive	LPF	Thresholding	Statistical Feature, Profile Feature	Pattern Matching
[61]	Wi-Fi	OFDM	Passive	STFT	Thresholding	Encode	Pattern Matching
[86]	Wi-Fi	Amplitude of CSI	Passive	Background Subtraction, PCA	Thresholding	Frequency Vector, Statistical Feature	HMM
[90]	RFID	Phase	Passive	Frequency Hops Normalization, Median Filter	Thresholding	Statistical Feature	Machine Learning
[27]	Wi-Fi, VL, mmWave, Ultrasonic	Amplitude	Passive	Hampel Filter	Time Window	CNN	DNN combined with discriminator
[99]	Fusion	Amplitude	Passive	Hampel Filter	Time Window	CNN	DNN Combined with Discriminator
[7]	Wi-Fi	Amplitude of CSI	Passive	Conjugate Multiplication, LPF, PCA, STFT	Time Window	Frequency Vector	HMM
[96]	Ultrasonic	RSS	Passive	STFT	Thresholding	PCA	Voting Mechanism, Random Forest, SVM
[79]	Wi-Fi	Amplitude of CSI	Passive	Conjugate Multiplication, STFT, Background Subtraction, PCA	Thresholding	Frequency Vector	Jaccard Similarity Coefficient
[94]	Wi-Fi	Amplitude	Passive	LPF	Morphology Matching	Statistical Feature	Voting Mechanism, SVM

Table 3. Performance of Activity Identification Systems

Reference	Performance Evaluation						Description
	Overall	Robustness	Stability	Generality	Multiuser	Scalability	
[113]	99.68%	92.75%	99.68%	-	1	1	Calibration-free motion detection
[37]	99.4%	99.42%	-	97%	1	1	Boundary-aware motion detection
[22]	94%	94%	-	-	1	1	Fall detection
[84]	93%	91.25%	79.75%	86.2%	1	1	Fall detection
[115]	86.83%	86.83%	-	-	1	1	Adaptable fall detection
[89]	>94%	>80%	-	-	1	9	Multi-activity classification
[61]	94%	94%	90%	-	4	9	Multi-activity classification
[86]	96.5%	>80%	-	-	1	9	Multi-activity classification
[90]	93.5%	>85.5%	>83%	-	1	8	Multi-activity classification
[27]	>50%	-	-	-	1	6	Environment-independent activity classification
[99]	87.9%	81.2%	-	84.2%	1	8	Multi-modal activity classification
[7]	91.3%	>74%	-	>83%	1	7	Driving activity classification
[96]	94.8%	-	-	>91.73%	1	8	Early driving activity classification
[79]	>90%	>88%	-	-	6	15	Multi-user activity classification
[94]	93.1%	<6% (ERR)	-	<2% (ERR)	3	7	Multi-user activity classification

detection, and *multi-activity classification*. The relevant works are summarized in Tables 2 and 3 based on the general research framework discussed in section 2.

Motion detection aims at detecting the presence of moving things. It is a classical binary classification problem of WDHS. Once persons enter a monitored place, their movements will change the propagation paths of the wireless signal and cause burst signal patterns at the receivers. Conventional methods [86] leverage pre-defined thresholds to distinguish such motion-induced burst signals. However, they need to be calibrated for different environments since wireless signals are sensitive to the environment and environmental changes. To overcome this problem, WiDetect [113] used statistical theory to model the signal states in scattering-rich environments. Though WiDetect is robust and calibration-free and has broad coverage, it still suffers from false alarms in real life due to the fuzzy sensing range. WiBorder [37] gave an in-depth analysis of CSI conjugate multiplication and proposed a sensing boundary determination method. Thus, it can precisely tell whether a person enters a given place. Nevertheless, the deployment of transceivers requires a specific design, which may influence the user experience of the fundamental communication function.

Fall detection is a typical problem of detecting a specific target in daily activities. It is non-trivial to find a feature that can clearly distinguish falls from other movements compared with motion detection. Therefore, WiFall [22] and RT-Fall [84] selected data-driven methods for fall detection, e.g., random forest and SVM. As we know, such data-driven models require big datasets to remove bias and improve generality. However, collecting data in different environments is laborious, and labeling wireless signals requires professional knowledge due to their unintuitive expression. The researchers of TL-Fall [115] observed that the knowledge learned in old environments had a positive influence on training models in new environments. Based on this observation, they introduced a feature-based transfer learning method to reduce the data collection costs in new environments.

Multi-activity classification is related to what the activity is, and the recognition systems need more features to distinguish multiple activities. E-eyes [89] directly compared the signal patterns by ***Dynamic Time Warping* (DTW)**. WiSee [61], CARM [86], and TACT [90] converted the wireless signals into the frequency domain to obtain the environment-independent DFS (indicating movement speed information) and leveraged machine learning models for classification. To improve the environmental robustness, EI [27] and DeepMV [99] combined DNN with a domain discriminator to extract the general features of activity. WiDrive [7] and ER [96] explored the methods to detect dangerous driving activities. In the scenarios of driving, the drivers are fixed to their seats, and the in-car environments are stable. Therefore, WiDrive and ER do not need to consider the robustness of environmental changes. The prior works are constrained to single-user scenarios. Distinguishing the in-home activities of multiple users at the same time is more challenging. WiMU [79] analyzed the channel frequency response power model of the combined movements of multiple users and designed a virtual sample generation method to reduce data collection costs. However, this method can only recognize the gestures performed, not who performed the gestures. Motion-Fi [94] leveraged the short interaction range of Wi-Fi-based backscatter tags to separate users physically. This method requires a specific design of deployment to avoid the overlapping of different tags' sensing areas.

Challenges and research opportunities: Most activity identification systems we reviewed are based on Wi-Fi since such devices are most prevalent in modern homes. These systems have limited practicability before the breakthrough in robustness and multiuser support. To improve the robustness, finding environment-independent representation is important. However, it is very challenging to derive the representation manually due to the sophisticated domain knowledge. Alternatively, researchers can benefit from deep learning techniques like EI [27] and DeepMV [99]. Then, there is the problem of reducing the costs of data collection. Designers can leverage few-shot learning techniques, e.g., transfer learning [59] and meta-learning [19]. Besides, they can choose data simulation techniques [5, 8] to generate virtual samples to accelerate the model training phase. Multiuser support is also important for home-level activity identification. Current Wi-Fi-based systems working at 2.4 GHz and 5 GHz cannot separate one person from another due to their narrow

Table 4. Methodology of Limb Motion Recognition Systems

Reference	Data Collection			Preprocessing	Segmentation	Representation	Sensing Model
	Source	Measurement	Type				
[29]	RFID	RSS	Passive	Moving Average, Background Subtraction	Thresholding	Profile Feature	*If-then-else* Statement
[80]	Wi-Fi	CSI	Passive	Background Subtraction, PCA	Thresholding, Preamble	Location Feature, Statistical Feature	kNN
[121]	Wi-Fi	CSI	Passive	Conjugate Multiplication, STFT, Background Subtraction	Time Window	BVP	*Recurrent Neural Network* (RNN) with GRU
[13]	Ambient RF	Noise Floor	Passive	Data Reconstruction, CSSD	Thresholding	Frequency Vector	Markov Chain Model
[65]	Wi-Fi	CSI	Passive	Conjugate Multiplication, BPF, PCA, STFT	Thresholding	Frequency Vector, LPF, STFT	Pattern Matching

Table 5. Performance of Limb Motion Recognition Systems

Reference	Performance Evaluation						Description
	Overall	Robustness	Stability	Generality	Multiuser	Scalability	
[29]	97%	-	-	-	1	8	Low-power arm gesture recognition
[80]	91.4%	-	>90%	-	1	6	Location-aware arm gesture recognition
[121]	92.7%	>87%	-	>85%	1	6	Location-aware arm gesture recognition
[13]	92.2%	-	-	-	1	8	Arm gesture recognition via ambient RF
[65]	92%	-	-	>85%	1	9	Leg motion direction recognition

bandwidth. In order to identify the activities of co-existing users, high bandwidth is necessary. Hence, we envision that more techniques with high bandwidths will be applied for multiuser activity identification.

3.2 Limb Motion Recognition

Limb motions are the basic forms to send instructions to intelligent devices. As summarized in Tables 4 and 5, the selected papers can be divided into two groups according to the target body parts: arm gesture recognition and leg gesture recognition.

Arm gesture recognition aims to recognize some common actions, e.g., push, pull, and slide, for interacting with smart devices. In the era of the ***Internet of Things*** **(IoT)**, many IoT devices with limited computational resources appear in our lives. One practical problem is the requirement of low power consumption. AllSee [29] used power-harvesting sensors to provide energy to RFID tags for gesture classification. It leveraged *if-then-else* statements to distinguish eight predefined *arm movements*. However, it was sensitive to position changes since a given movement induced different multipath patterns in different locations. To overcome this problem, WiAG [80] proposed a transfer function based on a theoretical analysis of the relationship between gestures in different locations. It can estimate user location through a simple preamble movement and recognize the following gestures by comparing them with virtual samples. WiDar3.0 [121] introduced the ***body-coordinate velocity profile*** **(BVP)**, integrating movement speed, orientation, and location, to realize location-independent gesture recognition.

Leg motion recognition has a narrow scope of application since legs are not as flexible as arms. WiDance [65] proposed a leg motion recognition method for interactive exergames where the moving direction was essential information. Inspired by the prior Wi-Fi-based human sensing work [61], WiDance analyzed the relations between motion direction and the related DFS. Then, it designed a specific deployment of two T-R pairs to recognize nine different leg gestures with different moving directions.

Table 6. Methodology of Hand Gesture Recognition Systems

Reference	Data Collection			Preprocessing	Segmentation	Representation	Sensing Model
	Source	Measurement	Type				
[1]	Wi-Fi	RSSI	Passive	DWT	Thresholding, Preamble	Encode	Pattern Matching
[36]	Wi-Fi	Amplitude of CSI	Passive	Outlier Removal, Moving Average, LPF	Thresholding	Profile Feature	DTW, KNN
[75]	Wi-Fi	Amplitude of CSI	Passive	DWT, Long Delay Removal	Thresholding	Weighted Profile Feature	DTW
[6]	Wi-Fi	CSI	Passive	LPF, PCA, Normalization	Dynamic Thresholding	DWT	DTW, KNN
[68]	Ultrasonic	DFS	Active	Background Subtraction, Gaussian Smoothing	Time Window	Radical Velocity	Pattern Matching
[11]	Ultrasonic	Amplitude	Active	STFT	Time Window	Encode	Pattern Matching
[40]	Ultrasonic	CIR	Active	First-order Difference, LPF	Thresholding	Image	CNN
[124]	RFID	Phase	Passive	Unwrapping, Savitzky-Golay Filter, Normalization	Thresholding	Profile Feature	Voting Mechanism, DTW
[116]	RFID	Phase	Passive	Unwrapping, Savitzky-Golay Filter, Normalization	Thresholding	Statistical Feature, Profile Feature	Random Forest
[45]	Solar	Photocurrent amplitude	Passive	DWT	Thresholding	Statistical Feature, Profile Feature	Machine Learning
[50]	Wi-Fi	CSI	Passive	LPF	Thresholding	Profile Feature	Cross-correlation, DTW
[69]	mmWave	RSS	Active	LPF, HPF, STFT	Preamble	CNN	Multitask Learning Network
[18]	Audio	Amplitude	Passive	Background Subtraction, STFT, LPF	Thresholding	Gray-scale	LeNet-5
[106]	Audio	Amplitude	Passive	Moving Average	Threshold	Statistical Feature, MFCC	SVM
[81]	RFID	CIR	Active	Interpolation, Moving Average	Time Window	Pearson Correlation Coefficient, RSSI Distribution	KNN, CNN

Challenges and research opportunities: Limb motion recognition has a stronger interaction tendency than activity identification. The systems are designed for room-level application scenarios where the users may interact with multiple IoT devices simultaneously. Besides the challenging requirement of low-power consumption, a practical limb motion recognition system should be context-aware and available for mobile persons. When IoT devices increase, it needs many instructions to control them. It is very difficult to map every instruction to a unique limb gesture. From the perspective of user experience, the design of instruction gestures should be simple and natural. For example, users usually use a finger swipe to switch videos playing on their phones. So, it is more natural to use an arm swipe to switch songs playing on smart speakers and shows on TV. Therefore, a practical system can infer which device the user wants to control by aggregating the information of the user's direction and location. In addition, limb-motion-induced signals are weaker than those caused by torso movements. Hence, existing systems require users to stand still when performing limb motions to avoid the interference of the torso. Designers can select the signals with a fine-grained spatial resolution to recognize limb behaviors in mobile scenarios.

3.3 Hand Gesture Recognition

Hand gesture recognition has attracted increasing attention since the potential target gestures can express as much information as the human voice. The application scenarios include *in-air hands-free input*, *ASL recognition*, and *handwriting recognition*, as summarized in Tables 6 and 7.

In-air hands-free input maps predefined hand or finger gestures to control instructions. Hand and finger gestures are smaller than limb motions, inducing much smaller wireless signal changes.

Table 7. Performance of Hand Gesture Recognition Systems

Reference	Performance Evaluation						Description
	Overall	Robustness	Stability	Generality	Multiuser	Scalability	
[1]	96%	96%	-	-	1	9	Wi-Fi-based hand gesture recognition
[36]	>90%	-	-	-	1	9	Wi-Fi-based finger gesture recognition
[75]	93%	90%	>90%	>90%	1	6	Wi-Fi-based finger gesture recognition
[6]	93.47%	-	-	-	1	36	Keystroke recognition
[68]	95.1%	>85%	-	>85%	1	6	Acoustic-based hand gesture recognition
[11]	96.8%	-	96.8%	-	1	6	Hand gesture recognition for file sharing
[40]	97.92%	85.67%	>88.81%	92.56%	1	6	Acoustic-based hand gesture recognition
[124]	96.5%	93.7%	>60%	-	1	6	RFID-based hand gesture recognition
[116]	>94%	-	-	-	1	9	Real-time hand gesture recognition
[45]	>96%	>85%	>94.2%	>85%	1	6	Solar-based hand gesture recognition
[50]	92%	88%	>67%	>72%	1	25	ASL recognition
[69]	87%	86.7%	78%	>75.2%	1	50	Long-range ASL recognition
[18]	88%	-	-	74.83%	1	26	Handwriting recognition
[106]	55%	-	-	64.94%	1	26	Privacy risk of handwriting recognition
[81]	>89%	-	-	>89%	1	26	Multi-touch handwriting recognition

Such small signal changes are usually buried in strong ones caused by other body parts or moving objects, and it is non-trivial to extract them. Therefore, researchers often constrain the interaction range to a small space to zoom into the signals of hands and fingers.

WiGest [1] required users to perform gestures close to laptops to capture the hand-induced RSSI changes of Wi-Fi signals. WiFinger [36] fixed the distance between transceivers at 50 *cm* and asked users to place their hands on the LOS path. It measured the CSI with higher spatial resolution than RSSI and leveraged KNN to classify nine finger gestures. Another work, also named WiFinger [75], achieved high robustness and resilience to individual diversity through a weighted profile comparison method. By assigning higher weights to the more representative parts, this work can reduce the negative impact of movement instability and user diversity. The authors of WiKey [6] observed that each keystroke had a unique CSI pattern due to the different hand gestures. Then, it used a threshold-based method for segmentation, DWT for feature extraction, and KNN for keystroke recognition, achieving high accuracy of 93.47%. However, this system was sensitive to the changes in the relative positions between users and devices.

In recent years, researchers have paid much attention to acoustic-based hand recognition methods based on the built-in speakers and microphones of IoT devices. AudioGest [68] turned the device into an active sonar. By analyzing the features extracted from the audio DFS, it can recognize six hand gestures to control media applications with an accuracy of up to 96% within a range of 25 *cm*. AirLink [11] leveraged the direction information of DFS and realized hand-controlled file sharing. However, these DFS-based methods had a low spatial resolution, limiting their performance on recognizing fine-grained gestures, especially involving fingers. Hence, UltraGesture [40] utilized ***Channel Impulse Response*** **(CIR)** with higher resolution. Then, it regarded the measurements as images and leveraged a ***convolutional neural network*** **(CNN)** for gesture recognition.

RFID-based gesture recognition is attracting increasing attention from academia and industry due to the properties of lightweight, low power consumption, and convenient deployment. GRfid [124] matched the signal segments to a set of registered templates. Such a pattern matching method had a low generalization ability and increased the recognition latency. ReActor [116] extracted coarse-grained and fine-grained features through statistic-based and DWT-based methods, respectively. Then, it leveraged random forest for more robust and real-time performance.

An interesting work, namely SolarGest [45], proposed a solar-based gesture recognition approach to control solar-powered smart devices. The key insight was that hand gestures near the transparent solar panel, transforming solar light to photocurrent signals, can induce unique photocurrent signal patterns. Hence, SolarGest analyzed the correlations between the hand gestures and the motion-induced photocurrent signals and designed an end-to-end machine-learning-based framework for recognition.

ASL recognition systems need to capture richer features to distinguish a wider range of hand gestures. Melgarejo et al. [50] explored the application scenarios of controlling a wheelchair or a car through ASL gestures. In such scenarios, the users were constrained in their seats, and their torsos and legs were static relative to their arms. Therefore, the ASL-induced signals would not be covered by the stronger signals caused by other body parts' movements. Moreover, the researchers leveraged directional antennas near the target hands to capture fine-grained features and selected the most similar class via DTW. Santhalingam et al. [69] proposed an mmWave-based ASL recognition system, extending the interaction range to 3 m away from the devices.

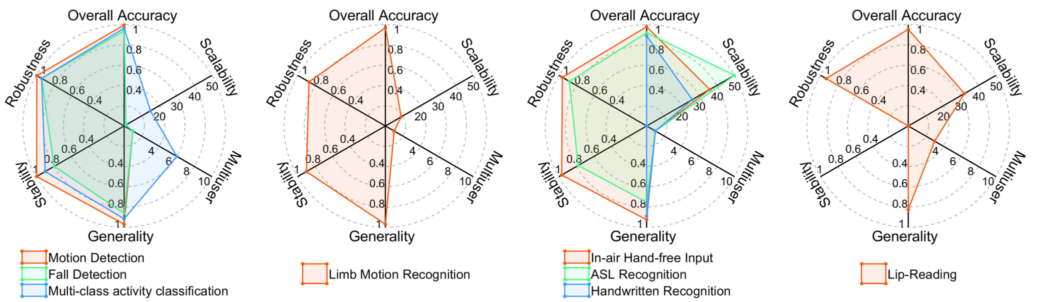
Handwriting recognition. Existing works of handwriting recognition mainly use acoustic signals from friction between pens and papers. WordRecorder [18] utilized microphones deployed near fingers to receive the audio signals propagating on the desktop for handwriting recognition. WritingHacker [106] explored the risk of eavesdropping on handwriting by a nearby microphone. It obtained stroke and letter segmentation through a threshold-based method and recognized words by SVM. RF-Finger [81] designed a 2D RFID tag array for in-air handwriting recognition. When a finger moves close to an RFID tag, the finger-reflected RSSI becomes stronger. Hence, RF-Finger regarded the 2D tag array's RSSI distribution as images and fed them into CNN for recognition.

Challenges and research opportunities: Compared with other body parts, hands can express the most meanings. To distinguish hand gestures accurately, one challenge is that hand-induced signals are much lower than other body parts. Existing systems usually deploy sensing devices near the hands. However, the application scenarios are limited. On account of this, researchers can leverage directional antennas and beamforming techniques to zoom in on the space of hands like mmASL [69] or explore more robust signal models like CSI-quotient [110]. The second challenge is to extract distinguishable features. Since it is relatively difficult to model hand gestures, we envision that more data-driven models will be adopted in the future.

3.4 Lip Reading

Wireless signals can not only "see" what people write but also "hear" what people say. WiHear [82] leveraged directional antennas to send Wi-Fi signals in the direction of the target user's face. Thus, the reflected signals contained the information of mouth motions, and WiHear introduced a DTW-based classification method to recognize 14 syllables and 32 words with an accuracy of 91%. This system required a user to stand in front of the devices, making it unsuitable for mobile scenarios. WiTalk [17] explored a more general application scenario of lip reading while making a phone call. The relative location between phone and mouth is stable no matter where the user is or in which direction the user faces. It measured the DFS to obtain the motion information of the mouth and adopted DTW to classify 12 syllables, achieving higher than 82.5% accuracy. SilentTalk [73] turned a mobile phone into a sonar to capture fine-grained motion features. Then, it introduced a probability model for lip reading and achieved an accuracy of 95.4%.

Challenges and research opportunities: Except for the limitations of sensing range due to the slight movements, the practicality of lip motion recognition is also constrained by the assumption of simple transition states, e.g., short pauses between words or syllables. Based on this assumption, current systems leveraged threshold-based methods to decompose the signals and narrow the target space. However, the transition states are complex in real life and closely related to their context. Therefore, we envision that more context-aware models will be applied for lip reading. For example, designers can refer to the methods for sequence-to-sequence problems such as Transformer [78].



(a) Whole-body (b) Limbs (c) Hands/Fingers (d) Lips

Fig. 3. Research status of behavior recognition application scenarios.

To show the research gaps of behavior recognition, we conclude the current research status of the involved eight applications with six metrics of overall accuracy, robustness, stability, generality, multi-user support, and scalability, as shown in Figure 3. The binary classification problems, including *motion detection* and *fall detection*, are close to practical requirements. However, there is still room for improvement, such as incorporating boundary detection to reduce the false-positive rate. A common problem for the home-level *multi-class activity classification* and *limb motion recognition* is multi-user support. Hands/fingers-related applications are well researched. Hands and fingers are more flexible than limbs but induce much smaller signal variations, resulting in a small interaction space. To a certain degree, small interaction spaces can filter out the environmental noise and reduce the impact of user diversity existing in other body parts. Therefore, *in-air hands-free input* and *ASL recognition* have good robustness, stability, and generality performance. *Handwriting recognition* lacks evaluation of robustness and stability, but we think it should have comparable performance because of the similar sensing task and granularity. Lip reading is an emerging research field, and there are many existing challenges of stability, generality, and multi-user support.

4 MOVEMENT TRACKING

This section summarized the WDHS systems that track human behaviors. According to the target sizes, the selected papers can be classified into four categories: *whole-body movement tracking*, *pose estimation*, *hands/fingers tracking*, and *vital sign monitoring*.

4.1 Whole-body Movement Tracking

Whole-body movement tracking systems mainly focus on three applications: *location sensing*, *crowd counting*, and *step counting*. Location sensing systems concentrate on the coordinates of the target people. Crowd counting systems estimate how many people are in an interesting place. The step counting systems can estimate the number of steps to assist in efficient walking based on the repeated gait-induced signal patterns. We summarize the methodology and performance of whole-body movement tracking systems in Tables 8 and 9, respectively.

Location sensing has attracted significant attention in the field of WDHS because it provides an opportunity to locate or track people without requiring them to carry on-body devices. It outperforms the traditional device-based methods in specific application scenarios, e.g., locating tourists in the scenic area to prevent them from entering dangerous areas or damaging the environment. Tadar [100] deployed an RFID tag array on the outer wall of a target room to track the indoor target. It extracted the motion-induced signals by subtracting the learned empty room information since the strong signals reflected by the wall would not be changed by people moving indoors. Widar [63]

Table 8. Methodology of Whole-body Movement Tracking Systems

Reference	Data Collection			Preprocessing	Segmentation	Representation	Sensing Model
	Source	Measurement	Type				
[100]	RFID	Amplitude and Phase	Active	Background Subtraction	Time Window	-	HMM
[63]	Wi-Fi	CSI	Passive	BPF, PCA, STFT, Moving Average	Time Window	Velocity Vector	Dead Reckoning
[64]	Wi-Fi	CSI	Passive	Conjugate Multiplication, HPF	Time Window	ToF, AoA, DFS	Mathematical Mapping
[3]	FM	FMCW	Active	Background Subtraction	Time Window	ToF	Ellipse Localization
[2]	FM	FMCW	Active	Background Subtraction	Time Window, Silhouette Cancellation	Heatmap	Peak Detection
[92]	Wi-Fi	Amplitude of CSI	Passive	Long Delay Removal	Time Window	PEM	Probabilistic Model
[33]	Wi-Fi	Phase of CSI	Passive	PCA, HPF	Time Window	CFPs, CSPs	Probabilistic Model
[102]	VL	Voltage	Passive	Moving Average	Time Window	Geographical Feature, Temporal Feature	Support Vector Regression
[98]	Wi-Fi	CSI	Passive	Long Delay Removal, BPF	Time Window	PCA, DWT, Short-time Energy	Peak Detection
[114]	Wi-Fi	Amplitude of CSI	Passive	Savitzky-Golay Filter	Time Window	Canonical Polyadic Decomposition	Peak Detection

Table 9. Performance of Whole-body Movement Tracking Systems

Reference	Performance Evaluation						Description
	Overall	Robustness	Stability	Generality	Multiuser	Scalability	
[100]	9.8 cm	-	-	-	1	<7.6 m	Through-wall location sensing
[63]	25 cm	-	-	<50 cm	1	<10 m	Location sensing
[64]	75 cm	<51 cm	<100 cm	<75 cm	1	<10 m	Location sensing with a single Wi-Fi link
[3]	<17.7 cm	<21 cm	-	-	1	<11 m	Location sensing via FMCW radar
[2]	<15.9 cm	<16.1 cm	-	-	4	<7 m	Through-wall multi-user location sensing
[92]	98%	84%	-	-	30	<8 m	Crowd counting
[33]	96.3%	94.23%	-	-	10	-	Crowd counting for stationary people
[102]	>92%	-	>90%	-	20	-	Crowd counting using existing LEDs
[98]	>87.6%	88.9%	-	>80%	1	<8 m	Step counting
[114]	>84.53%	-	-	-	5	-	Multi-runner step counting

and Widar2.0 [64] tracked users in 2D coordinates via two orthogonal Wi-Fi links relying on the relationship between motion velocity and the relevant DFS. At the beginning of wireless sensing, scientists developed radars to track moving objects. Inspired by this idea, Adib et al. proposed a ***frequency modulated carrier wave*** **(FMCW)**-based tracking system, WiTrack [3], working in the frequency band of WLAN. It measured the ToF of FMCW and estimated the propagation distance. Then, WiTrack leveraged the ellipse model to track a single person. In a further study, WiTrack2.0 [2] designed a subtraction strategy to locate multiple users. After locating one target, it removed the corresponding impulses from the measurements and then found the next target iteratively.

Crowd counting, focusing on the exact number of people in a specific place, is highly important in smart living. For example, when people gather in dangerous places, it can automatically notify security officers to avoid man-made disasters. The typical strategy is to find a metric with a monotonic relationship with the number of people. The device-free crowd counting system, FCC [92], introduced a metric, ***percentage of nonzero elements*** **(PEM)**, formulating a monotonic relationship between CSI variation and the number of moving people. However, with the number of people increasing, PEM almost stopped growing, which caused FCC's performance degradation. The authors of CelingSee [102] observed that the diffuse reflection with respect to the number of people

Table 10. Methodology of Pose Estimation Systems

Reference	Data Collection			Preprocessing	Segmentation	Representation	Sensing Model
	Source	Measurement	Type				
[119]	FM	FMCW Symbol	Active	Background Subtraction	Time Window	Heatmap	TSN
[120]	FM	FMCW Symbol	Active	Background Subtraction	Time Window	Heatmap	CNN
[28]	Wi-Fi	CSI	Passive	Conjugated Multiplication, STFT	Time Window	Velocity Vector	CNN-RNN
[117]	EMF	RSS	Passive	HPF	Time Window	Z-Score	Threshold
[38]	VLC	RSS	Passive	STFT	Thresholding	Shadow Map	Optimization, Greedy Algorithm

had a monotonic relationship with the lightning-induced voltage changes in the conventional bidirectional interface between the LED and its micro-controller. Korany and Mostofi [33] mapped the counting stationary people to the M/G/∞ queuing problem. They introduced two specific metrics: ***crowd fidgeting periods*** **(CFPs)**, indicating the periods in which at least one person had small in-place natural body movements, and ***crowd silent periods*** **(CSPs)**, indicating the periods in which no person moved. Then, Korany et al. utilized a thresholding-based motion detection method to obtain such metrics and then leverage the ***maximum a posteriori*** **(MAP)** to estimate the number of people based on the M/G/∞ queuing theory.

Step counting. It is challenging to track steps directly because the corresponding signals of legs or feet might be covered by the stronger ones reflected by the torso. Based on the insight that the torso and legs move at different speeds and cause different frequency shifts, WiStep [98] utilized time-frequency analysis to recognize the repeated walking patterns. It leveraged DWT to separate the leg-induced signals from the torso-induced ones and a peak detection method to count steps. WiStep achieved a step counting accuracy higher than 87.6% in 2D space. However, it required a single person in the monitoring space. Wi-Run [114] proposed a multi-user-supported step counting method and applied it for workout assistance. It deployed a transceiver and a receiver on both sides of a set of treadmills. The users were asked to run on the treadmills. Wi-Run assumed that running-induced CSI measurements of a single person can be approximated by sinusoid-like patterns. Then, a tensor decomposition function was implemented to separate the running-related signals of multiple users. Hence, Wi-Run realized counting the steps of up to five users simultaneously and achieved an average accuracy higher than 84.5%.

Challenges and research opportunities: The whole-body movement tracking systems have simple sensing targets, such as people's coordinates [63], people numbers [33], and step numbers [98]. Since tracking whole-body movements is a coarse-grained sensing task, the sensing devices can be deployed far from the targets. However, a larger sensing area suffers from more frequent interference from other moving objects. Hence, robustness and multi-user support are two important properties. To improve the performance on these aspects, designers can improve the sensing systems' spatial resolution by increasing the number of antennas [2] or using signals with higher bandwidths [69].

4.2 Pose Estimation

Limb-oriented tracking systems concentrate on the locations of people's joints, namely *pose estimation*. Pose estimation is well studied in the computer vision community but is an emerging area in WDHS. Wireless signals are waves that cannot intuitively show the spatial relationship of joints like images. Fortunately, in recent years, wireless techniques have made a breakthrough in this area, as shown in Tables 10 and 11.

Table 11. Performance of Pose Estimation Systems

Reference	Performance Evaluation						Description
	Overall	Robustness	Stability	Generality	Multiuser	Scalability	
[119]	93.3%	>85%	-	-	14	<12 m	FM-based pose estimation via TSN
[120]	<4.9 cm	<5.2 cm	-	-	5	<10 m	Multi-user 3D pose estimation
[28]	3.67 cm	5.42 cm	7.91 cm	8.58 cm	1	-	Pose estimation via COTS Wi-Fi devices
[117]	8.6 cm	-	-	-	1	<3 m	Pose estimation by transforming a wall to a touchscreen
[38]	10°	-	<10°	<16°	1	-	Skeleton reconstruction

It is promising to search with the help of other mature techniques, e.g., image-based pose estimation [119, 120] and VICON [28]. Such systems deployed cameras or VICON systems in parallel with their wireless sensing devices to simultaneously record human behaviors. Then, they trained pose estimation models in a semi-supervised learning framework such as the ***teacher-student network* (TSN)**. The experimental results showed that these systems can achieve comparable performance to their teacher systems. Researchers have explored other techniques for pose estimation without the assistance of external systems. Inspired by the modern touchscreens, Wall++ [117] turned the common wall into a giant touchscreen to sense human motion in a near place based on EMF signals. Based on the insight that shadows are 2D projects of the objects that block light beams, LiSense [38] built up a photodiode matrix to obtain coarse-grained shadow information. Then, it regarded the skeleton reconstruction as an optimization problem and selected a greedy algorithm to speed up the pose estimation.

Challenges and research opportunities: Wireless-based pose estimation is a challenging task because of the difficulty of modeling human postures. The wireless signals are not intuitive enough to show the spatial correlations between joints. Thus, most systems take advantage of DNN to learn the joints' spatial knowledge from wireless signals under the guidance of other techniques, such as image and VICON. Compared with the teacher systems, such sensing models have comparable performance in LOS conditions and much higher accuracies in NLOS conditions. The common limitation of existing systems is the cost of data collection because they need to deploy additional devices to collect other signals in parallel. To ensure proper guidance, the device of the teacher system needs to be set up in the same location as the student system and kept in sync. Therefore, it would be troublesome to improve the robustness of the system by changing the layout to collect data from different perspectives. To reduce the data collection costs, we envision that more studies will leverage simulation technology to generate virtual data by changing the device places in the digital world [5].

4.3 Hand/Finger Tracking

Hand and finger movement tracking systems monitor the moving trajectories of hands or fingers. The goals of these systems are to extend the interaction ways of intelligent devices with small screens. Generally, they involve three applications: *hand tracking*, *finger tracking*, and *keystroke tracking*, as shown in Table 12.

Hand tracking shares a similar idea to location sensing systems. It aims at estimating hand positions in real time. EchoTrack [9] turned a mobile device into sonar and tracked hands through simple mathematical mapping algorithms. WiDraw [72] proposed a hand tracking method via Wi-Fi signals. It regarded hands as moving transmitters and introduced the typical ***multiple signal classification* (MUSIC)** algorithm to obtain the AoA measurements. To meet the requirement of precisely controlling an intelligent device, e.g., adjusting the volume of sounds, QGesture [105] quantified hand movements using ***commercial off-the-shelf* (COTS)** Wi-Fi devices. It calculated the propagation path changes via the observation on the hand-induced DFS and leveraged a triangulation model to quantify the hand movements.

Table 12. Methodology of Hand/Finger Tracking Systems

Reference	Data Collection			Preprocessing	Segmentation	Representation	Sensing Model
	Source	Measurement	Type				
[9]	Ultrasonic	CIR	Active	BPF, Long Delay Removal	Time Window	ToF	Mathematical Mapping
[72]	Wi-Fi	CSI	Passive	MUSIC, LPF	Time Window	Azimuth, Elevation, Depth	Mathematical Mapping
[105]	Wi-Fi	CSI	Passive	Interpolation, Moving Average, PCA	Preamble	Phase-distance Relationship	Triangulation Model
[54]	Ultrasonic	OFDM Symbol	Active	STFT	Time Window	Phase-distance Relationship	Ellipse Model
[87]	Ultrasonic	Phase	Active	Local Min-max Averaging, Cascaded Integrator Comb Filter	Time Window	Phase-distance Relationship	Ellipse Model
[108]	Ultrasonic	CIR	Active	Background Subtraction	Time Window	Phase-distance Relationship	Ellipse Model
[111]	VL	RSS	Active	Background Subtraction	Time Window	RSS Vectors	Pattern Matching
[39]	VL	RSS	Active	STFT	Thresholding	Shadow Map	Quasi-random Search
[91]	Wi-Fi	CSI quotient	Passive	Savizky-Golay Filter	Threshold	Phase-distance Relationship	Ellipse Model
[122]	Audio	RSS	Passive	-	Thresholding	TDoA	Hyperbola Model
[43]	Audio	RSS	Passive	-	Thresholding	TDoA, MFCC	K-means

Finger tracking systems usually track fine-grained finger motions in an active manner. They design specific modulated wireless signals and leverage directional antennas tracking finger motions like radar. Thus, smart devices can track fingers through typical ellipse models [54, 87, 108]. Okuli [111] proposed a VL-based system tracking a single finger in a 2D plane through fingerprinting. According to the reflection coefficient modeling how the finger bends the propagation paths, Okuli generated a lookup table consisting of the virtual RSS fingerprints. Then, it located fingers through a simple pattern matching method. Aili [39] designed a table lamp with photodiodes to obtain fine-grained shadow profiles of hands within 54 *cm*. Through an optimization search algorithm, Aili tracked not only hand location but also hand posture. To further improve the tracking accuracy of Wi-Fi, the passive tracking system FingerDraw [91] proposed a CSI-quotient model. It can cancel random phase offsets and remove uncertain impulse noise in the received CSI data. As a result, CSI-quotient maximized the SNR and decreased the passive finger tracking error to 12.7 *mm*.

Keystroke tracking infers the inputs of keystrokes according to finger trajectories, exposing the privacy concern of keystroke-based eavesdropping. Since the finger tracking systems can locate the coordinates of fingers in a 2D plane, researchers can reconstruct keyboards based on the possibility of the keystroke location. Zhu et al. [122] estimated the TDoA of keystroke sound propagating to different microphones on one phone. Every TDoA can determine the region of a keystroke. The overlapping regions from multi-TDoA can precisely locate the position of the keystroke. In another work, Liu et al. [43] used only one phone to obtain a set of candidate keystrokes by TDoA. Then, they analyzed the Mel-Frequency Cepstral Coefficients to select the accurate one from the candidate keystrokes.

Challenges and research opportunities: Hand/finger tracking has the potential to provide a natural interaction way with smart devices, especially in the near future of Metaverse. As summarized in Table 13, most of the existing systems focus on the coordinates' information and do not track hand postures. Hence, people cannot interact with virtual objects through natural motions such as grabbing. Tracking hand posture is relatively challenging because the hands are too flexible to be modeled, and fingers are so thin that it is non-trivial to separate their relevant signals. To

Table 13. Performance of Hand/Finger Tracking Systems

Reference	Performance Evaluation						Description
	Overall	Robustness	Stability	Generality	Multiuser	Scalability	
[9]	<3.6 cm	-	<4 cm	-	1	<80 cm	Hand tracking via a COTS smartphone
[72]	<5.4 cm	<6 cm	<5.76 cm	-	1	<60 cm	Hand tracking via COTS Wi-Fi devices
[105]	<3.75 cm	<5 cm	<5 cm	<4 cm	1	<200 cm	Quantifying hand motion
[54]	8 mm	<13.5 mm	-	<10 mm	1	<50 cm	Finger tracking via smartwatches and smartphones
[87]	4.57 mm	-	5.81 mm	-	1	<20 cm	Finger tracking
[108]	3 mm	<13.5 mm	-	-	1	<20 cm	Finger tracking
[111]	<7.5 mm	<7.5 mm	<7.5 mm	-	1	<13 cm	VL-based finger tracking
[39]	2.5 mm	-	-	-	1	<54 cm	Hand pose reconstruction
[91]	12.7 mm	<20 mm	-	<20 mm	1	<100 cm	Finger tracking via CSI-quotient
[122]	72.2%	-	>64%	-	1	<10 cm	Keystroke tracking
[43]	>85.5%	-	-	-	1	<13 cm	Keystroke tracking via single smartphone

Table 14. Methodology of Vital Sign Monitoring Systems

Reference	Data Collection			Preprocessing	Segmentation	Representation	Sensing Model
	Source	Measurement	Type				
[83]	Wi-Fi	Phase of CSI	Passive	Hampel Filter, Moving Average	Time Window, Thresholding	Power Spectral Density	Peak Detection
[110]	Wi-Fi	CSI quotient	Passive	Savitzky-Golay Filter	Time Window	Profile Feature	Autocorrelation, Peak Detection
[12]	RFID	RSS, Phase	Passive	Moving Average, BPF	Time Window	Power Spectral Density	Peak Detection
[57]	FM	Phase, RSS	Active	LPF, Background Subtraction	Time Window	Profile Feature	Neural Network
[97]	Ultrasonic	ESD	Active	Background Subtraction, BPF, STFT	Time Window	Helbert Spectrum	GAN
[4]	FM	FMCW Symbol	Active	Background Subtraction	Time Window, Thresholding	Profile Feature, Frequency Vector	Pattern Matching, Peak Detection
[88]	Wi-Fi	Phase of CSI	Passive	Hampel Filter, DWT	Time Window	Profile Feature, Frequency Vector	root-MUSIC, Peak Detection, Pattern Matching
[62]	Ultrasonic	FMCW Symbol	Active	IIR comb notch filter	Time Window	Frequency Vector	Peak Detection

Table 15. Performance of Vital Sign Monitoring Systems

Reference	Performance Evaluation						Description
	Overall	Robustness	Stability	Generality	Multiuser	Scalability	
[83]	-	-	-	-	1	<6 m	Fresnel-based respiration monitoring
[110]	<0.5 bpm	<0.34 bpm	<0.3 bpm	-	1	<8 m	Long-range respiration monitoring
[12]	>93%	-	>90%	98.7%	2	-	Multi-user respiration monitoring
[57]	>90%	-	-	-	1	-	Respiration waveform monitoring
[97]	0.11 bpm	-	<0.22 bpm	0.11 bpm	1	<0.3 m	Respiration waveform monitoring
[4]	99%	<90.1%	<91.7%	-	3	<8 m	Respiration and heartbeat monitoring
[88]	<1 bpm	<0.52 bpm	-	-	4	-	Respiration and heartbeat monitoring
[62]	19 ms	-	-	<50 ms	1	<0.3 m	Heartbeat monitoring by smartphone

improve tracking hand posture, one can use signals with higher spatial resolution [39] and deep learning techniques.

4.4 Vital Sign Monitoring

It is challenging for device-free systems to monitor respiration and heartbeats due to the subtle chest movements. Tables 14 and 15 show that the typical applications of WDHS are *respiration monitoring* and *heartbeat monitoring*. Some vital sign monitoring systems use ***beats per minute* (bpm)** to describe heart rate according to clinical experience.

Respiration monitoring systems often leverage peak detection to count the respiration rate based on the repeated patterns of the chest movements. Wang et al. [83] designed a Fresnel-Zone-based model to monitor human respiration without training, and it was robust to different positions and orientations. FarSense [110] leveraged the CSI-quotient model to expand the monitoring range to 8 meters. These systems were designed for the situation of a single user. However, a respiration monitoring system should simultaneously monitor multiple users in the real application scenario, e.g., hospitals. Thus, LungTrack [12] proposed an RFID-based multi-user respiration monitoring method. For each propagation link between readers and tags, there existed dead zones that had poor sensing performance. Inspired by this observation, LungTrack carefully designed the deployment of the tags to ensure that dead zones can separate every user, thus realizing monitoring multiple people simultaneously. Another problem limiting the practice was monitoring moving targets. WiSpiro [57] proposed an FM-based respiration system designed for sleeping people. When a user changed his/her sleeping pose, WiSpiro automatically moved the sensing devices to the proper place where they could capture clear chest movements. Then, a neural network was implemented for non-contact respiration monitoring. BreathListener [97] extended the acoustic-based vital sign monitoring to a dynamic scenario of the driving environment. It extracted the ***energy spectrum density*** **(ESD)** of the received acoustic data. Then, this system utilized a ***generative adversarial network*** **(GAN)** to generate fine-grained respiration waveforms.

Heartbeat monitoring systems have a common problem: heartbeat-induced vibrations are orders of magnitude lower than breath-induced chest movements. Thus, the heartbeat-related signals are normally masked by chest movements. Vital-Radio [4] observed that the rates of heartbeats were usually in the range from $40\,Hz$ to $200\,Hz$, while the respiration rate was lower than $20\,Hz$. Hence, it can filter the heartbeat-induced signal from the frequency perspective. Similarly, PhaseBeat [88] leveraged DWT for movement separation and monitored respiration and heartbeat at the same time, reaching accuracies of 0.5 *bpm* and 1 *bpm*, respectively. ***Acousticcardiogram*** **(ACG)** [62] transformed the commercial smartphone into an FMCW sonar and monitored spatial changes of the chest. Then, it utilized an IIR comb notch filter to obtain heartbeat-induced signals.

Challenges and research opportunities: Device-free vital sign monitoring has a promising application future since it does not bring additional wearing burden to users like traditional methods. With the increase of sensing precision, we envision that more studies will explore monitoring more precise indicators, e.g., heartbeat waveform and lung functions [71]. Further, one can monitor other health indices related to respiration and heartbeats. For example, TagSleep [41] proposed a low-cost solution for sleep state identification, including snore, cough, and somniloquy, through respiration-induced chest movements. Based on the assumption that some physiological signals like heartbeats change with emotion, EQ-Radio [118] extracted the heartbeat from wireless signals and leveraged SVM to identify human emotion, achieving an average accuracy of 87% of four emotions, e.g., joy, sadness, anger, and neutral. In addition, we observe that even for the same sensing task, researchers may select different metrics [12, 110] to evaluate the performance. This will make it difficult for readers know which one is better. For a complete comparison, a generally accepted metric selection rule is required.

Finally, we provide an overview of the current state of movement tracking applications. As shown in Figure 4, the scalability of current model-based crowd detection systems has reached the upper bound, counting up to 30 people within 8 meters. When the number of co-existing people grows, the signal indicators selected are nearly unchanged. Hence, researchers need more powerful feature extraction methods to break this limitation. The step counting systems lack evaluation to validate their long-term working performance, and multi-user support is also a problem. We put the three applications, e.g., location sensing, pose estimation, and hand tracking, in one group since they select the same accuracy evaluation metric. At present, WDHS-based location sensing

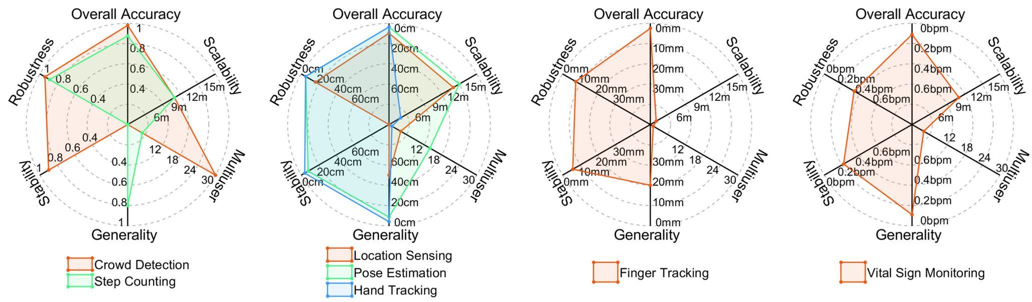


Fig. 4. Research status of movement tracking application scenarios.

Table 16. Methodology of Whole-body Motion Authentication Systems

Reference	Data Collection			Preprocessing	Segmentation	Representation	Sensing Model
	Source	Measurement	Type				
[109]	Wi-Fi	CSI	Passive	Long Delay Removal, BPF	Thresholding, Peak-valley Detection	Statistical Feature	Decision Tree
[85]	Wi-Fi	CSI	Passive	Background Subtraction, LPF, PCA, STFT	Thresholding, Autocorrelation	Profile Feature, Spectrogram	SVM
[123]	Wi-Fi	Amplitude of CSI	Passive	DWT	Peak-valley Detection	Profile Feature	Shapelet Learning
[95]	Ultrasonic	RSS	Active	Band-stop Filter, BPF, STFT	Thresholding	Statistical Feature	SVM
[51]	mmWave	Point Cloud	Active	DBSCAN, Hungarian Algorithm	Time Window	Horizontal Angle, Pitching Angle of Points, Distance	CNN
[25]	RFID	Phase, RSSI	Passive	Interpolation, Unwrapping	Time Window	-	CNN

systems cannot be implemented in public places with more than 10 co-existing people. Hand tracking application achieves a good performance within an interaction range shorter than 2 meters. Thanks to the specific devices, pose estimation can track more than 10 people simultaneously at a distance of around 12 meters. Finger tracking aims at broadening the ways of interacting with intelligent devices at a close distance. Hence, multi-user support is not necessary. However, there is still a big gap to be bridged for practical applications such as the impact of user diversity. Vital sign monitoring achieved accurate tracking of the respiration and heartbeat of up to four people within a range of 8 meters. For more practical use, researchers and developers can explore the way of tracking finer-grained signs like electrocardiograms.

5 USER IDENTIFICATION

User identification systems leverage behavioral features to distinguish different people. Each person has unique behavioral patterns due to the user diversity caused by different body shapes, gaits, and health conditions. For gesture classification models, user diversity may harm the generalization ability of new users. However, on the other hand, it provides an opportunity to identify users. According to the motion granularity, the existing wireless-based user identification systems can be grouped into three categories: *whole-body-motion-based authentication*, *finger-motion-based authentication*, and *lip-motion-based authentication*.

5.1 Whole-body-motion-based Authentication

In Tables 16 and 17, we summarize the methodology of whole-body motion authentication systems and their performance. Existing systems mainly use the personal characteristics in gait and daily

Table 17. Performance of Whole-body Motion Authentication Systems

Reference	Performance Evaluation						Description
	Overall	Robustness	Stability	Generality	Multiuser	Scalability	
[109]	>80%	-	-	-	1	6	Gait recognition via COTS Wi-Fi devices
[85]	>79.28%	-	-	-	1	50	Gait recognition via COTS Wi-Fi devices
[123]	91%	-	-	-	1	20	Gait recognition via shapelet learning
[95]	96.6%	-	-	>86.5%	1	50	Gait recognition via a mobile phone
[51]	88%	-	>86%	-	5	50	Multi-user gait recognition
[25]	97.72%	-	-	-	5	15	User authentication via daily activities

activities for authentication. Therefore, the systems can be classified into *gait recognition* and *daily-activity-based authentication*.

Gait recognition provides a silent authentication solution when target users pass through a specific sensing area. Since gait-induced signals have repeated patterns, most gait recognition systems conform to a common strategy. They extract the repeated gait cycle from the signal streams and then leverage machine learning [85, 95, 109] or DNN [123] for user identification. mmGait [51] used FMCW radars working at 60GHz to obtain the 3D point cloud of walking people. Then, it extracted the location, the radial speed, and the signal speed of the points to train a DNN-based user identifier. The model achieved an average identification accuracy of 90% and 88% for single-user and multi-user scenarios, respectively.

Daily-activity-based authentication. Continuous authentication plays an essential role in future smart home scenarios [15, 48]. It can protect privacy and prevent families from dangerous conditions. For example, continuous authentication can ensure that a smart oven will not respond to naughty kids' orders. One problem with continuous authentication systems is how to make them transparent. Based on the assumption that the personal feature exists in not only a specific motion type but also our daily activities, Au-Id [25] proposed an RFID-based *authentication method through sequential daily activities*. It deployed RFID tags on the infrastructure like doors with which people will interact. Based on the correlation between the tags and the infrastructure, Au-Id stacked a CNN with a ***long short-term memory*** **(LSTM)** to label different activities automatically. Then, it input the labeled data to another CNN for user identification and achieved an average identification accuracy of 97.72%.

Challenges and research opportunities: Unlike biometrics, such as fingerprints and irises, which remain constant, behavioral patterns can change with mental and health conditions. Therefore, behavior-based authentication methods require regular updates to improve their stability. However, in real life, it is impossible to ask every person to provide a set of training data to fine-tune the authentication models when they enter a monitored place. One can take advantage of another modality of authentication methods to ease the pain of data collection. For example, in an early attempt XModal-ID [32] combined video-based authentication with Wi-Fi-based authentication to track unknown people in unknown places. We envision that cross-modal authentication techniques will be implemented to improve stability.

5.2 Finger-motion-based Authentication

With the development of IoT, people will frequently interact with intelligent appliances where finger motions play an important role. Based on the user diversity in finger motions, FingerPass [31] proposed continuous finger-motion-based authentication methods via Wi-Fi signals. It proposed a three-layer LSTM model to extract features of gesture, motion, and user, respectively. This model can simultaneously recognize finger motions and identify who performed them with an authentication accuracy of 91.4%.

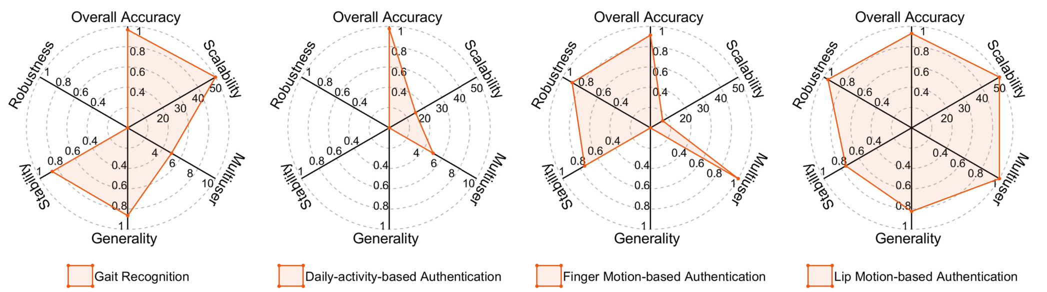


Fig. 5. Research status of user identification application scenarios.

Challenges and research opportunities: FingerPass and Au-ID [25] (mentioned in Section 5.1) expose the fact that user diversity widely exists in human behaviors. These works validated the feasibility of behavioral authentication by conducting small-scale experiments. However, the result cannot answer how many users can be identified through a certain finger motion. It lacked a theory or model to describe the feasibility of behavior-based user authentication. Since it is relatively challenging to find such a theory, we envision that more daily activities will be used for authentication, and a theory or a model will be proposed to guide the system design.

5.3 Lip-motion-based Authentication

Lips have many features for user identification, e.g., lip-prints, lip shape, and lip color [14]. However, these features require high-resolution techniques like images. At present, wireless-based sensing techniques cannot meet the resolution requirement. Hence, lip-oriented wireless user identification systems distinguish people based on lip motions.

Since lip motion contains personal characteristics, SilentKey [74] obtained the rhythm (i.e., the interval time between two consecutive mouth motions) and the duration (i.e., the time spent for inputting passwords) from the received CIR signals. It utilized SVM to estimate users and achieved 70% to 83.1% user ID accuracy, and the spoofer detection accuracy ranged from 86.7% to 90.7%. LipPass [44] designed a three-layer autoencoder network to extract the lip motion features and SVM to identify the users. It achieved 90.2% user identification accuracy and 93.1% spoofer detection accuracy.

Challenges and research opportunities: Lip motion is a soft biometric for user identification. It may change with human health conditions. Therefore, lip-motion-based user identification systems should be adaptable to keep their performance stable over a long-term period. A straightforward way is to build up big datasets and train a general model. However, it requires tedious efforts of data collection. To reduce such costs, one can design adaptable models such as transfer learning [59] or online learning.

In Figure 5, we show the Research status of user identification. WDHS-based gait recognition systems are deployed in some important corridors of a smart building for authentication, and their deployment seldom changes. In such scenarios, robustness to environmental changes is not a necessary indicator. More important in practical use is multi-user support, as it is common to have multiple people in public corridors simultaneously. Leveraging daily activities and finger motions for user identification are two emerging research fields, and there are many challenges that need to be solved. WDHS lip-motion-based authentication has comparable performance to mature commercial applications based on face recognition techniques [44]. However, the related systems should improve the stability to meet the requirement of long-term usage.

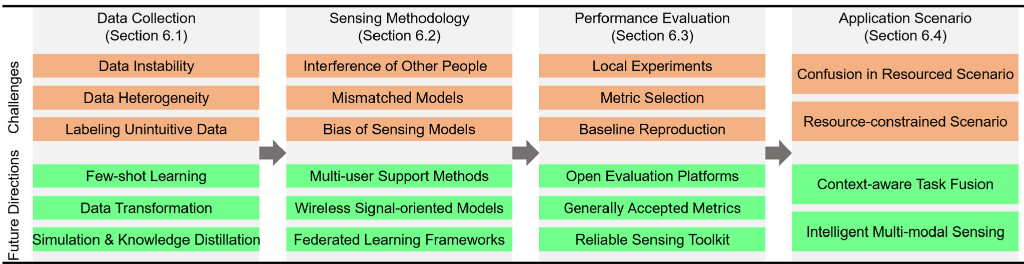


Fig. 6. Challenges and future research directions of WDHS.

6 CHALLENGES & RESEARCH DIRECTIONS

This section concludes with 11 challenges from the aspects of *data collection*, *sensing methodology*, *performance evaluation*, and *application scenario*. In addition, we provide corresponding research directions for each challenge, as shown in Figure 6.

6.1 Data Collection

Challenges: In recent years, there was a trend of leveraging experience-based models to sense complex movements or activities. To train robust sensing models is relatively challenging due to the following issues: (1) *Data instability*. Unlike the constant biological characteristics such as fingerprints, iris, and retina, human behaviors change over time, so the induced wireless signals will also change over time. As a result, the performance of a human sensing system may degrade when it works for a long time. (2) *Data heterogeneity*. Researchers have explored a wide range of wireless sensing techniques, e.g., Wi-Fi, acoustic, mmWave, and Image. Due to the differences in devices, deployment design, and system parameters, many datasets may be built for similar sensing targets from different views. The related models are not general, resulting in limited practicality. (3) *Labeling unintuitive data*. The received wireless signals are in waves and are not as intuitive as images. Thus, it is difficult for humans to distinguish different movements only according to the wireless signals, making it impossible to offload data annotation tasks to remote people.

Research directions: To address these challenges, we suggest that one can benefit from the techniques of *few-shot learning*, *data transformation*, and *simulation and knowledge distillation*.

Few-shot learning. In the past decade, few-shot learning techniques were well studied to address the lack of data. To ease the pain of data instability caused by collection costs, one can leverage active learning [67] that selects the most representative unknown samples for annotation. Since people's behavior changes are not sudden, we can use online learning [23] to gradually adjust sensing systems. Alternatively, researchers can leverage transfer learning [59, 93] and meta-learning [19] techniques to reduce the data collection efforts.

Data transformation. Designers can build bridges between heterogeneous data through data transformation methods. The heterogeneous data provides multiple views describing the same human movements. Based on this observation, some attempts [5, 8] converted heterogeneous data into the same feature space to realize cross-modal WDHS systems. Specifically, they first construct a human body mesh from videos via image processing techniques [30]. Through analyzing the vertex location changes of the body mesh, they can obtain the velocity of each body part and simulate the DFS measurements derived from propagation models. Such methods reveal a promising research direction of pre-trained WDHS models. People can train such models on massive public videos. Since the videos involve many people and movements, the pre-trained WDHS models will have better performance on robustness than the existing wireless sensing models. Moreover, the pre-trained models can quickly adapt to new modality techniques if we convert the data into speed-related feature spaces.

Simulation and knowledge distillation. To ease the pain of labeling unintuitive data, one can leverage simulation techniques to generate virtual samples of human behaviors. For example, we can leverage game engines, reinforcement learning techniques [60], or GAN [20] to simulate the 3D motions of human behavior and map the 3D motions to the time-frequency space based on path loss models. Thus, it is possible to obtain countless virtual samples to train a pre-trained model and only need a small volume of labeled data for fine-tuning in real applications. In addition, researchers can learn from the idea of knowledge distillation to reduce costs of data annotation like the prior attempts [28, 119, 120]. They leverage non-wireless-based systems to train a WDHS model in semi-supervised learning frameworks, and the wireless sensing models can achieve comparable performance to the teacher systems.

6.2 Sensing Methodology

Challenges: (1) *Interference of other people.* Most WDHS systems have limited practicality because they only work well in single-user conditions. They cannot separate the people moving around, resulting in performance degradation. (2) *Mismatched models.* With the development of deep learning, DNN shows an excellent ability to learn the representation of complex objects. Since current DNN models are designed for image-based sensing tasks, the input data's format requirement is structured and non-complex numbers. Existing systems must convert the wireless signals to an image-like format before feeding them to the models [40]. This signal conversion operation may result in the loss of some information. (3) *Bias of sensing models.* Existing sensing models are trained on local datasets covering limited environments and users. As a result, these models exhibit bias on the training dataset and poor generalization ability.

Research directions: For the challenges of interference of other people, mismatched models, and bias of sensing models, we envision the following research directions:

Multi-user support methods are a necessary property for many applications, e.g., activity identification and location sensing. To sense the behaviors of co-existing people by COTS devices is relatively challenging. For example, WiMU [79] leveraged COTS Wi-Fi devices to recognize the combined gestures of multiple users. However, they cannot build the connections between user and gesture. MotionFi [94] and LungTrack [12] took advantage of the tags' short sensing range to sense multiple users, but they needed a specific deployment design. It required a theoretical model or experimental study to discuss the boundaries of COTS-based multi-user sensing. Except for the costs of devices, specific hardware-based methods [2, 51] have better accuracy and a more convenient deployment strategy. We envision that more techniques with high spatial resolution will be implemented for multi-user sensing.

Wireless-signal-oriented models. Wireless signals have unique features such as frequency and phases that reflect human movements. These unique features provide additional views to represent human movements. However, it is very challenging to leverage these unique features directly. In terms of taking advantage of frequency, Yao et al. [103] proposed STFNets, combining DNN with time-frequency analysis to learn the motion representation in the frequency domain. Experimental results showed that STFNets outperformed the baselines of DNN-based wireless sensing models. It validated that the unique features can improve the performance of WDHS systems. Hence, we envision that more sensing models specifically designed for wireless signals will be proposed.

Federated learning frameworks. The straightforward way to mitigate bias is by expanding the training datasets, covering more people, gestures, and environments. However, people may not like to share their data to the public due to privacy concerns. A promising research direction to overcome the bias may be the combination of *Federated Learning* [101] and WDHS. Intuitively, federated learning is a secure way to transfer knowledge. It can train models on local datasets without directly sharing the data. Therefore, we can "gather" all local datasets safely and cooperatively train general models.

6.3 Performance Evaluation

Challenges: Existing WDHS systems with the same sensing target cannot be fully compared because of the following problems: (1) *Local experiments*. Existing works evaluate their performance on local testing datasets collected from different users in different environments. These varied factors also make their performance lack comparability. (2) *Metric selection*. Different systems may select incomparable metrics for the same sensing target to evaluate their performance. For example, FarSense [110] selected *bpm* error to evaluate the accuracy of respiration monitoring, while LungTrack [12] defined the accuracy as a ratio between the correctly detected number and the ground truth. (3) *Baseline reproduction*. Since the model parameters are not described in detail in the original work, researchers may get different results when they try to reproduce it. Therefore, the comparison results may be unconvincing.

Research directions: To improve the comparability, researchers can provide guidance in the aspects of *open evaluation platform*, *generally accepted metrics*, and *reliable sensing toolkit*.

Open evaluation platform. Researchers have made efforts to overcome the problem of local experiments. For example, Yousefi et al. [104] proposed a public dataset involving six people and six activities. Ma et al. [47] constructed a sign language dataset including 276 sign words. Zheng et al. [121] published the Widar3.0 dataset with 258,575 gestures referring to 22 types provided by 17 users in 75 domains. However, the in-total volume of WDHS open datasets is not large enough compared with other research communities' datasets like ImageNet [26]. It is a long-term effort for WDHS researchers to build a big open evaluation platform.

Generally accepted metrics. Systems for different applications have different requirements. For example, location sensing systems use location error to evaluate their accuracy, and multi-user support is a potentially important metric. For lip-motion-based authentication systems, they may select F1-Score to evaluate accuracy. In their application scenarios, users will interact with intelligent devices at a short distance, so multi-user support is unnecessary. Therefore, the WDHS community needs studies to tell us whether a WDHS system can meet the practical requirement of real application scenarios. Also, we need guidance on which properties are essential in a specific application scenario and how we conduct experiments to validate such properties.

Reliable sensing toolkit. Besides the open platform and metric guidance, it would be better if there was a toolkit to help reproduce the baselines correctly. Due to differences in understanding, people may have deviations when reproducing a certain job, which reduces the credibility of the comparison results.

6.4 Application Scenario

Challenges: From the perspective of the application scenario, WDHS encounters different challenges of *confusion in resourced scenarios* and *resource-constrained environments*:

(1) *Confusion in resourced scenarios*. With IoT devices increasing, WDHS systems need to avoid confusing users' instructions. For example, when a user performs a "light up" motion in the living room, he does not want to find the kitchen light. For another case, a smart home should distinguish the gestures from different people for personalized services.

(2) *Resource-constrained scenarios*. There will be some blind spots that the sensing system cannot cover in resource-constrained environments, e.g., construction sites. Existing WDHS systems require costly methods of expanding the device topology since they are designed for environments where sufficient signal resources are used.

Research directions: These challenges can be addressed by *context-aware task fusion* and *intelligent multi-modal sensing*.

Context-aware task fusion develops systems with multiple sensing capabilities to meet the requirements in different application scenarios, making them more practical to complex sensing tasks. For example, multiTrack [76] proposed a Wi-Fi-based location-aware human sensing method, integrating behavior recognition with location sensing. It can recognize different behaviors of coexisting users. Guo et al. [21] fused activity identification with user identification and proposed a personal assistance system providing custom workout recommendations. EZ-Sleep [24] fused location sensing with vital sign monitoring, and it automatically tracked the users' sleep states without knowing the beds' locations.

Intelligent multi-modal sensing. We recommend that researchers explore the multi-modal sensing solutions from two perspectives: fusing unknown ambient signals and fusing known signals.

The key problem with fusing unknown ambient signals is signal instability. Since most ambient signals are sent by sensors, the sparsity and instability make them more difficult for human sensing. The sensing systems, e.g., EAR [13], should use as many wireless signals as possible. However, the more wireless signal sources a system uses, the more unstable are the received signal states. In addition, the transmitters are invisible, and they may be far from the target person. Also, the signal propagation paths are easily interfered with by moving objects or people. Furthermore, it needs more complex signal processing tools or sensing models to extract the motion-induced signals.

For fusing known signals, "When to fuse?" is an important question that needs to be carefully considered. On the one side, the signal sources in poor conditions may negatively influence the sensing tasks. On the other side, it requires a flexible power management strategy since more devices engage in a single sensing task, resulting in higher power consumption. Therefore, the forthcoming multi-modal sensing systems must adjust their fusion schemes according to the environmental context. Based on the general framework of WDHS systems, different sensing techniques can be fused at three levels: *data level*, *feature level*, and *result level*.

- *Data level* means that a multi-modal system fuses the heterogeneous data at the beginning to unify the input format. Through combining with other modal data, researchers can mitigate the inherent shortages of wireless signals, e.g., instability. For example, human gait is unstable due to changes in health conditions. Wireless-based gait recognition systems require long-term data collection efforts to maintain good performance. To overcome this problem, one can leverage the mature image-based authentication method to give footage of the wireless data when an unknown person first appears [32]. At the same time, the system can implicitly fine-tune its wireless-based gait recognition model for continuous authentication.
- *Feature level* transforms different data into the same feature space. For example, DeepMV [99] utilized CNN to extract features from received Wi-Fi and ultrasonic measurements and generate feature vectors of the same size. Then, it fuses the vectors as the input of a DNN-based sensing model. The main advantage is that the sensing model can learn the fusion weights automatically during the training process. As a result, the feature-level fusing strategy can avoid the negative influence of some signal types with poor sensing performance according to the environmental conditions.
- *Result level* means that all the candidate techniques are parallel, and their outputs will be fused. This strategy is simple, modular, and flexible. However, researchers should have an in-depth understanding of the candidate systems, making sure that their outputs are useful to the sensing task or should be ignored. Thus, they can avoid the fused results from being misled by some untrustworthy outputs.

7 CONCLUSION

This survey gives an up-to-date review of *wireless device-free human sensing* (WDHS) systems from the perspectives of sensing task type and motion granularity. We grouped the WDHS systems into 12 categories according to their sensing task type and granularity. We summarized related research under a general research framework to expose researchers and developers with an overview of the typical methodology and current performance. Finally, the article discussed the challenges from *data collection*, *sensing methodology*, *performance evaluation*, and *application scenario*, aiming to stimulate further efforts in WDHS.

Correspondingly, we envision that WDHS systems can reduce data collection costs by leveraging few-shot learning, data transformation, simulation, and knowledge distillation techniques. The sensing models will be more practical if they could sense co-existing people, match wireless signals, and be trained on various datasets under federated learning frameworks. With the help of generally accepted metric selection rules, reliable sensing toolkit, and open evaluation platform, the forthcoming WDHS systems can be accurately compared. As a result, developers will easily figure out the proper sensing methods. Finally, we believe that more context-aware multi-modal systems will be proposed to handle the complex sensing tasks in real life.

REFERENCES

[1] Heba Abdelnasser, Moustafa Youssef, and Khaled A. Harras. 2015. Wigest: A ubiquitous wifi-based gesture recognition system. In *Proceedings of the 34th Annual IEEE International Conference on Computer Communications (INFOCOM'15)*. IEEE, 1472–1480.

[2] Fadel Adib, Zachary Kabelac, and Dina Katabi. 2015. Multi-person localization via RF body reflections. In *Proceedings of the 12th USENIX Symposium on Networked Systems Design and Implementation (NSDI'15)*. USENIX, 279–292.

[3] Fadel Adib, Zach Kabelac, Dina Katabi, and Robert C. Miller. 2014. 3D tracking via body radio reflections. In *Proceedings of the 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI'14)*. USENIX, 317–329.

[4] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C. Miller. 2015. Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI'15)*. ACM, 837–846.

[5] Karan Ahuja and Chris Harrison. 2021. Vid2Doppler: Synthesizing Doppler radar data from videos for training privacy-preserving activity recognition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'21)*. ACM, 1–10.

[6] Kamran Ali, Alex Xiao Liu, Wei Wang, and Muhammad Shahzad. 2015. Keystroke recognition using wifi signals. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (MobiCom'15)*. ACM, 90–102.

[7] Yunhao Bai, Zejiang Wang, Kuangyu Zheng, Xiaorui Wang, and Junmin Wang. 2019. WiDrive: Adaptive WiFi-based recognition of driver activity for real-time and safe takeover. In *Proceedings of the 39th International Conference on Distributed Computing Systems (ICDCS'19)*. IEEE, 901–911.

[8] Hong Cai, Belal Korany, Chitra R. Karanam, and Yasamin Mostofi. 2020. Teaching RF to sense without RF training measurements. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 4, 4 (2020), 1–22.

[9] Huijie Chen, Fan Li, and Yu Wang. 2017. EchoTrack: Acoustic device-free hand tracking on smart phones. In *Proceedings of the 36th Annual IEEE International Conference on Computer Communications (INFOCOM'17)*. IEEE, 1–9.

[10] Kaixuan Chen, Dalin Zhang, Lina Yao, Bin Guo, Zhiwen Yu, and Yunhao Liu. 2021. Deep learning for sensor-based human activity recognition. *ACM Computing Surveys (CSUR)* 54, 4 (2021), 1–40.

[11] Ke-Yu Chen, Daniel Ashbrook, Mayank Goel, Sung-Hyuck Lee, and Shwetak Patel. 2014. AirLink: Sharing files between multiple devices using in-air gestures. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'14)*. ACM, 565–569.

[12] Lili Chen, Jie Xiong, Xiaojiang Chen, Sunghoon Ivan Lee, Daqing Zhang, Tao Yan, and Dingyi Fang. 2019. LungTrack: Towards contactless and zero dead-zone respiration monitoring with commodity RFIDs. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 3, 3 (2019), 1–22.

[13] Zicheng Chi, Yao Yao, Tiantian Xie, Xin Liu, Zhichuan Huang, Wei Wang, and Ting Zhu. 2018. EAR: Exploiting uncontrollable ambient RF signals in heterogeneous networks for gesture recognition. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems (SenSys'18)*. ACM, 237–249.

[14] Michał Choraś. 2010. The lip as a biometric. *Pattern Analysis and Applications (PAA)* 13, 1 (2010), 105–112.

[15] M. Jamal Deen. 2015. Information and communications technologies for elderly ubiquitous healthcare in a smart home. *Personal and Ubiquitous Computing (Special Issue on Aspects of Ubiquitous Computing for Improved Clinical Practice)* 19, 3–4 (2015), 573–599.

[16] Samundra Deep, Xi Zheng, Chandan Karmakar, Dongjin Yu, Leonard G. C. Hamey, and Jiong Jin. 2020. A survey on anomalous behavior detection for elderly care using dense-sensing networks. *IEEE Communications Surveys and Tutorials (COMST)* 22, 1 (2020), 352–370.

[17] Changlai Du, Xiaoqun Yuan, Wenjing Lou, and Y. Thomas Hou. 2018. Context-free fine-grained motion sensing using WiFi. In *Proceedings of the 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON'18)*. IEEE, 1–9.

[18] Haishi Du, Ping Li, Hao Zhou, Wei Gong, Gan Luo, and Panlong Yang. 2018. Wordrecorder: Accurate acoustic-based handwriting recognition using deep learning. In *Proceedings of the 37th Annual IEEE International Conference on Computer Communications (INFOCOM'18)*. IEEE, 1448–1456.

[19] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML'17)*. PMLR, 1126–1135.

[20] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Proceedings of the 28th International Conference on Neural Information Processing Systems (NeuIPS'14)*. MIT Press, 27.

[21] Xiaonan Guo, Jian Liu, Cong Shi, Hongbo Liu, Yingying Chen, and Mooi Choo Chuah. 2018. Device-free personalized fitness assistant using WiFi. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 2, 4 (2018), 1–23.

[22] Chunmei Han, Kaishun Wu, Yuxi Wang, and Lionel M. Ni. 2014. WiFall: Device-free fall detection by wireless networks. In *Proceedings of the 33th Annual IEEE International Conference on Computer Communications (INFOCOM'14)*. IEEE, 581–594.

[23] Steven C. H. Hoi, Doyen Sahoo, Jing Lu, and Peilin Zhao. 2021. Online learning: A comprehensive survey. *Neurocomputing* 459 (2021), 249–289.

[24] Chen-Yu Hsu, Aayush Ahuja, Shichao Yue, Rumen Hristov, Zachary Kabelac, and Dina Katabi. 2017. Zero-effort in-home sleep and insomnia monitoring using radio signals. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 1, 3 (2017), 1–18.

[25] Anna Huang, Dong Wang, Run Zhao, and Qian Zhang. 2019. Au-Id: Automatic user identification and authentication through the motions captured from sequential human activities using RFID. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 3, 2 (2019), 1–26.

[26] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'09)*. IEEE, 248–255.

[27] Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shuochao Yao, Yaqing Wang, Ye Yuan, Hongfei Xue, Chen Song, Xin Ma, Dimitrios Koutsonikolas, Wenyao Xu, and Lu Su. 2018. Towards environment independent device free human activity recognition. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking (MobiCom'18)*. ACM, 289–304.

[28] Wenjun Jiang, Hongfei Xue, Chenglin Miao, Shiyang Wang, Sen Lin, Chong Tian, Srinivasan Murali, Haochen Hu, Zhi Sun, and Lu Su. 2020. Towards 3D human pose construction using wifi. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking (MobiCom'20)*. ACM, 295–308.

[29] Bryce Kellogg, Vamsi Talla, and Shyamnath Gollakota. 2014. Bringing gesture recognition to all devices. In *Proceedings of the 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI'14)*. USENIX, 303–316.

[30] Muhammed Kocabas, Nikos Athanasiou, and Michael J. Black. 2020. Vibe: Video inference for human body pose and shape estimation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'20)*. IEEE, 5252–5262.

[31] Hao Kong, Li Lu, Jiadi Yu, Yingying Chen, Linghe Kong, and Minglu Li. 2019. FingerPass: Finger gesture-based continuous user authentication for smart homes using commodity WiFi. In *Proceedings of the 20th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'19)*. ACM, 201–210.

[32] Belal Korany, Chitra R. Karanam, Hong Cai, and Yasamin Mostofi. 2019. XModal-ID: Using WiFi for through-wall person identification from candidate video footage. In *Proceedings of the 25th Annual International Conference on Mobile Computing and Networking (MobiCom'19)*. ACM, 1–15.

[33] Belal Korany and Yasamin Mostofi. 2021. Counting a stationary crowd using off-the-shelf WiFi. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys'21)*. ACM, 202–214.

[34] Sylvia T. Kouyoumdjieva, Peter Danielis, and Gunnar Karlsson. 2020. Survey of non-image-based approaches for counting people. *IEEE Communications Surveys and Tutorials (COMST)* 22, 2 (2020), 1305–1336.

[35] Chenning Li, Zhichao Cao, and Yunhao Liu. 2021. Deep AI enabled ubiquitous wireless sensing. *ACM Computing Surveys (CSUR)* 54, 2 (2021), 1–35.

[36] Hong Li, Wei Yang, Jianxin Wang, Yang Xu, and Liusheng Huang. 2016. WiFinger: Talk to your smart devices with finger-grained gesture. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'16)*. ACM, 250–261.

[37] Shengjie Li, Zhaopeng Liu, Yue Zhang, Qin Lv, Xiaopeng Niu, Leye Wang, and Daqing Zhang. 2020. WiBorder: Precise Wi-Fi based boundary sensing via through-wall discrimination. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 4, 3 (2020), 1–30.

[38] Tianxing Li, Chuankai An, Zhao Tian, Andrew T. Campbell, and Xia Zhou. 2015. Human sensing using visible light communication. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (MobiCom'15)*. ACM, 331–344.

[39] Tianxing Li, Xi Xiong, Yifei Xie, George Hito, Xing-Dong Yang, and Xia Zhou. 2017. Reconstructing hand poses using visible light. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 1, 3 (2017), 1–20.

[40] Kang Ling, Haipeng Dai, Yuntang Liu, and Alex X. Liu. 2018. UltraGesture: Fine-grained gesture sensing and recognition. In *Proceedings of the 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON'18)*. IEEE, 1–9.

[41] Chen Liu, Jie Xiong, Lin Cai, Lin Feng, Xiaojiang Chen, and Dingyi Fang. 2019. Beyond respiration: Contactless sleep sound-activity recognition using RF signals. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 3, 3 (2019), 1–22.

[42] Jian Liu, Hongbo Liu, Yingying Chen, Yan Wang, and Chen Wang. 2020. Wireless sensing for human activity: A survey. *IEEE Communications Surveys and Tutorials (COMST)* 22, 3 (2020), 1629–1645.

[43] Jian Liu, Yan Wang, Gorkem Kar, Yingying Chen, Jie Yang, and Marco Gruteser. 2015. Snooping keystrokes with mm-level audio ranging on a single phone. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (MobiCom'15)*. ACM, 142–154.

[44] Li Lu, Jiadi Yu, Yingying Chen, Hongbo Liu, Yanmin Zhu, Yunfei Liu, and Minglu Li. 2018. LipPass: Lip reading-based user authentication on smartphones leveraging acoustic signals. In *Proceedings of the 37th Annual IEEE International Conference on Computer Communications (INFOCOM'18)*. IEEE, 1466–1474.

[45] Dong Ma, Guohao Lan, Mahbub Hassan, Wen Hu, Mushfika B. Upama, Ashraf Uddin, and Moustafa Youssef. 2019. SolarGest: Ubiquitous and battery-free gesture recognition using solar cells. In *Proceedings of the 25th Annual International Conference on Mobile Computing and Networking (MobiCom'19)*. ACM, 1–15.

[46] Yongsen Ma, Gang Zhou, and Shuangquan Wang. 2019. WiFi sensing with channel state information: A survey. *ACM Computing Surveys (CSUR)* 52, 3 (2019), 1–36.

[47] Yongsen Ma, Gang Zhou, Shuangquan Wang, Hongyang Zhao, and Woosub Jung. 2018. SignFi: Sign language recognition using WiFi. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 2, 1 (2018), 1–21.

[48] Sumit Majumder, Emad Aghayi, Moein Noferesti, Hamidreza Memarzadeh-Tehran, Tapas Mondal, Zhibo Pang, and M. Jamal Deen. 2017. Smart homes for elderly healthcare – Recent advances and research challenges. *Sensors* 17, 11 (2017), 2496–2527.

[49] Sumit Majumder, Tapas Mondal, and M. Jamal Deen. 2017. Wearable sensors for remote health monitoring. *Sensors* 17, 1 (2017), 45.

[50] Pedro Melgarejo, Xinyu Zhang, Parameswaran Ramanathan, and David Chu. 2014. Leveraging directional antenna capabilities for fine-grained gesture recognition. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'14)*. ACM, 541–551.

[51] Zhen Meng, Song Fu, Jie Yan, Hongyuan Liang, Anfu Zhou, Shilin Zhu, Huadong Ma, Jianhua Liu, and Ning Yang. 2020. Gait recognition for co-existing multiple people using millimeter wave sensing. In *Proceedings of the 2020 AAAI Conference on Artificial Intelligence (AAAI'20)*. AAAI, 849–856.

[52] Sushmita Mitra and Tinku Acharya. 2007. Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37, 3 (2007), 311–324.

[53] Athira Nambiar, Alexandre Bernardino, and Jacinto C. Nascimento. 2019. Gait-based person re-identification: A survey. *ACM Computing Surveys (CSUR)* 52, 2 (2019), 1–34.

[54] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. 2016. FingerIO: using active sonar for fine-grained finger tracking. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI'16)*. ACM, 1515–1525.

[55] Ebrahim Nemati, M. Jamal Deen, and Tapas Mondal. 2012. A wireless wearable ECG sensor for long-term applications. *IEEE Communications Magazine (Special Issue on Communications in Ubiquitous Healthcare)* 50, 1 (2012), 36–43.

[56] Kan Ngamakeur, Sira Yongchareon, Jian Yu, and Saeed Ur Rehman. 2020. A survey on device-free indoor localization and tracking in the multi-resident environment. *ACM Computing Surveys (CSUR)* 53, 4 (2020), 1–29.

[57] Phuc Nguyen, Xinyu Zhang, Ann Halbower, and Tam Vu. 2016. Continuous and fine-grained breathing volume monitoring from afar using wireless signals. In *Proceedings of the 35th Annual IEEE International Conference on Computer Communications (INFOCOM'16)*. IEEE, 1–9.

[58] Isura Nirmal, Abdelwahed Khamis, Mahbub Hassan, Wen Hu, and Xiaoqing Zhu. 2021. Deep learning for radio-based human sensing: Recent advances and future directions. *IEEE Communications Surveys and Tutorials (COMST)* 23, 2 (2021), 995–1019.

[59] Sinno Jialin Pan and Qiang Yang. 2010. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering (TKDE)* 22, 10 (2010), 1345–1359.

[60] Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel van de Panne. 2018. Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 1–14.

[61] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. 2013. Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th Annual International Conference on Mobile Computing and Networking (MobiCom'13)*. ACM, 27–38.

[62] Kun Qian, Chenshu Wu, Fu Xiao, Yue Zheng, Yi Zhang, Zheng Yang, and Yunhao Liu. 2018. Acousticcardiogram: Monitoring heartbeats using acoustic signals on smart devices. In *Proceedings of the 37th Annual IEEE International Conference on Computer Communications (INFOCOM'18)*. IEEE, 1574–1582.

[63] Kun Qian, Chenshu Wu, Zheng Yang, Yunhao Liu, and Kyle Jamieson. 2017. Widar: Decimeter-level passive tracking via velocity monitoring with commodity Wi-Fi. In *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'17)*. ACM, 1–10.

[64] Kun Qian, Chenshu Wu, Yi Zhang, Guidong Zhang, Zheng Yang, and Yunhao Liu. 2018. Widar2.0: Passive human tracking with a single wi-fi link. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys'18)*. ACM, 350–361.

[65] Kun Qian, Chenshu Wu, Zimu Zhou, Yue Zheng, Zheng Yang, and Yunhao Liu. 2017. Inferring motion direction using commodity Wi-Fi for interactive exergames. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI'17)*. ACM, 1961–1972.

[66] Adrian Reetz and Carl Gutwin. 2014. Making big gestures: Effects of gesture size on observability and identification for co-located group awareness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'14)*. ACM, 4087–4096.

[67] Pengzhen Ren, Yun Xiao, Xiaojun Chang, Po-Yao Huang, Zhihui Li, Brij B. Gupta, Xiaojiang Chen, and Xin Wang. 2021. A survey of deep active learning. *ACM Computing Surveys (CSUR)* 54, 9 (2021), 1–40.

[68] Wenjie Ruan, Quan Z. Sheng, Lei Yang, Tao Gu, Peipei Xu, and Longfei Shangguan. 2016. AudioGest: Enabling fine-grained hand gesture detection by decoding echo signal. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'16)*. ACM, 474–485.

[69] Panneer Selvam Santhalingam, Al Amin Hosain, Ding Zhang, Parth Pathak, Huzefa Rangwala, and Raja Kushalnagar. 2020. mmASL: Environment-independent ASL gesture recognition using 60 GHz millimeter-wave signals. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 4, 1 (2020), 1–30.

[70] Rathin Chandra Shit, Suraj Sharma, Deepak Puthal, Philip James, Biswajeet Pradhan, Aad van Moorsel, Albert Y. Zomaya, and Rajiv Ranjan. 2019. Ubiquitous localization (UbiLoc): A survey and taxonomy on device free localization for smart world. *IEEE Communications Surveys and Tutorials (COMST)* 21, 4 (2019), 3532–3564.

[71] Xingzhe Song, Boyuan Yang, Ge Yang, Ruirong Chen, Erick Forno, Wei Chen, and Wei Gao. 2020. SpiroSonic: Monitoring human lung function via acoustic sensing on commodity smartphones. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking (MobiCom'20)*. ACM, 1–14.

[72] Li Sun, Souvik Sen, Dimitrios Koutsonikolas, and Kyu-Han Kim. 2015. WiDraw: Enabling hands-free drawing in the air on commodity Wifi devices. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (MobiCom'15)*. ACM, 77–89.

[73] Jiayao Tan, Cam-Tu Nguyen, and Xiaoliang Wang. 2017. SilentTalk: Lip reading through ultrasonic sensing on mobile phones. In *Proceedings of the 36th Annual IEEE International Conference on Computer Communications (INFOCOM'17)*. IEEE, 1–9.

[74] Jiayao Tan, Xiaoliang Wang, Cam-Tu Nguyen, and Yu Shi. 2018. SilentKey: A new authentication framework through ultrasonic-based lip reading. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 2, 1 (2018), 1–18.

[75] Sheng Tan and Jie Yang. 2016. WiFinger: Leveraging commodity WiFi for fine-grained finger gesture recognition. In *Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'16)*. ACM, 201–210.

[76] Sheng Tan, Linghan Zhang, Zi Wang, and Jie Yang. 2019. MultiTrack: Multi-user tracking and activity recognition using commodity WiFi. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI'19)*. ACM, 1–12.

[77] Alexander Toshev and Christian Szegedy. 2014. DeepPose: Human pose estimation via deep neural networks. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'14)*. IEEE, 1653–1660.

[78] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeuIPS'17)*. MIT Press, 6000–6010.

[79] Raghav H. Venkatnarayan, Griffin Page, and Muhammad Shahzad. 2018. Multi-user gesture recognition using WiFi. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys'18)*. ACM, 1242–1256.

[80] Aditya Virmani and Muhammad Shahzad. 2017. Position and orientation agnostic gesture recognition using WiFi. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys'17)*. ACM, 252–264.

[81] Chuyu Wang, Jian Liu, Yingying Chen, Hongbo Liu, Lei Xie, Wei Wang, Bingbing He, and Sanglu Lu. 2018. Multi-touch in the air: Device-free finger tracking and gesture recognition via COTS RFID. In *Proceedings of the 37th Annual IEEE International Conference on Computer Communications (INFOCOM'18)*. IEEE, 1691–1699.

[82] Guanhua Wang, Yongpan Zou, Zimu Zhou, Kaishun Wu, and Lionel M. Ni. 2014. We can hear you with Wi-Fi! In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking (MobiCom'14)*. ACM, 593.

[83] Hao Wang, Daqing Zhang, Junyi Ma, Yasha Wang, Yuxiang Wang, Dan Wu, Tao Gu, and Bing Xie. 2016. Human respiration detection with commodity WiFi devices: Do user location and body orientation matter? In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'16)*. ACM, 25–36.

[84] Hao Wang, Daqing Zhang, Yasha Wang, Junyi Ma, Yuxiang Wang, and Shengjie Li. 2017. RT-Fall: A real-time and contactless fall detection system with commodity WiFi devices. *IEEE Transactions on Mobile Computing (TMC)* 16, 2 (2017), 511–526.

[85] Wei Wang, Alex X. Liu, and Muhammad Shahzad. 2016. Gait recognition using WiFi signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'16)*. ACM, 363–373.

[86] Wei Wang, Alex X. Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. 2015. Understanding and modeling of WiFi signal based human activity recognition. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (MobiCom'15)*. ACM, 65–76.

[87] Wei Wang, Alex X. Liu, and Ke Sun. 2016. Device-free gesture tracking using acoustic signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking (MobiCom'16)*. ACM, 82–94.

[88] Xuyu Wang, Chao Yang, and Shiwen Mao. 2017. PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices. In *Proceedings of the 37th International Conference on Distributed Computing Systems (ICDCS'17)*. IEEE, 1230–1239.

[89] Yan Wang, Jian Liu, Yingying Chen, Marco Gruteser, Jie Yang, and Hongbo Liu. 2014. E-eyes: Device-free location-oriented activity identification using fine-grained wifi signatures. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking (MobiCom'14)*. ACM, 617–628.

[90] Yanwen Wang and Yuanqing Zheng. 2018. Modeling RFID signal reflection for contact-free activity recognition. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 2, 4 (2018), 1–22.

[91] Dan Wu, Ruiyang Gao, Youwei Zeng, Jinyi Liu, Leye Wang, Tao Gu, and Daqing Zhang. 2020. FingerDraw: Sub-wavelength level finger motion tracking with WiFi signals. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 4, 1 (2020), 1–27.

[92] Wei Xi, Jizhong Zhao, Xiang-Yang Li, Kun Zhao, Shaojie Tang, Xue Liu, and Zhiping Jiang. 2014. Electronic frog eye: Counting crowd using WiFi. In *Proceedings of the 33th Annual IEEE International Conference on Computer Communications (INFOCOM'14)*. IEEE, 361–369.

[93] Jiang Xiao, Huichuwu Li, and Hai Jin. 2022. Transtrack: Online meta-transfer learning and Otsu segmentation enabled wireless gesture tracking. *Pattern Recognition (PR)* 121 (2022), 108157.

[94] Ning Xiao, Panlong Yang, Yubo Yan, Hao Zhou, and Xiang-Yang Li. 2018. Motion-Fi: Recognizing and counting repetitive motions with passive wireless backscattering. In *Proceedings of the 37th Annual IEEE International Conference on Computer Communications (INFOCOM'18)*. IEEE, 2024–2032.

[95] Wei Xu, Zhiwen Yu, Zhu Wang, Bin Guo, and Qi Han. 2019. AcousticID: Gait-based human identification using acoustic signal. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 3, 3 (2019), 1–25.

[96] Xiangyu Xu, Hang Gao, Jiadi Yu, Yingying Chen, Yanmin Zhu, Guangtao Xue, and Minglu Li. 2017. ER: Early recognition of inattentive driving leveraging audio devices on smartphones. In *Proceedings of the 36th Annual IEEE International Conference on Computer Communications (INFOCOM'17)*. IEEE, 1–9.

[97] Xiangyu Xu, Jiadi Yu, Yingying Chen, Yanmin Zhu, Linghe Kong, and Minglu Li. 2019. BreathListener: Fine-grained breathing monitoring in driving environments utilizing acoustic signals. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys'19)*. ACM, 54–66.

[98] Yang Xu, Wei Yang, Jianxin Wang, Xing Zhou, Hong Li, and Liusheng Huang. 2018. WiStep: Device-free step counting with WiFi signals. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 1, 4 (2018), 1–23.

[99] Hongfei Xue, Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shiyang Wang, Ye Yuan, Shuochao Yao, Aidong Zhang, and Lu Su. 2020. DeepMV: Multi-view deep learning for device-free human activity recognition. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 4, 1 (2020), 1–26.

[100] Lei Yang, Qiongzheng Lin, Xiangyang Li, Tianci Liu, and Yunhao Liu. 2015. See through walls with COTS RFID system! In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (MobiCom'15)*. ACM, 487–499.

[101] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. 2019. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)* 10, 2 (2019), 1–19.

[102] Yanbing Yang, Jie Hao, Jun Luo, and Sinno Jialin Pan. 2017. CeilingSee: Device-free occupancy inference through lighting infrastructure based LED sensing. In *Proceedings of the 15th Annual IEEE International Conference on Pervasive Computing and Communications (PerCom'17)*. IEEE, 247–256.

[103] Shuochao Yao, Ailing Piao, Wenjun Jiang, Yiran Zhao, Huajie Shao, Shengzhong Liu, Dongxin Liu, Jinyang Li, Tianshi Wang, Shaohan Hu, Lu Su, Jiawei Han, and Tarek Abdelzaher. 2019. Stfnets: Learning sensing signals from the time-frequency perspective with short-time fourier neural networks. In *Proceedings of the 30th Anniversary of the World Wide Web Conference (WWW'19)*. ACM, 2192–2202.

[104] Siamak Yousefi, Hirokazu Narui, Sankalp Dayal, Stefano Ermon, and Shahrokh Valaee. 2017. A survey on behavior recognition using WiFi channel state information. *IEEE Communications Magazine* 55, 10 (2017), 98–104.

[105] Nan Yu, Wei Wang, Alex X. Liu, and Lingtao Kong. 2018. QGesture: Quantifying gesture distance and direction with WiFi signals. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 2, 1 (2018), 1–23.

[106] Tuo Yu, Haiming Jin, and Klara Nahrstedt. 2016. Writinghacker: Audio based eavesdropping of handwriting via mobile devices. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'16)*. ACM, 463–473.

[107] Zhiwen Yu, He Du, Fei Yi, Zhu Wang, and Bin Guo. 2019. Ten scientific problems in human behavior understanding. *CCF Transactions on Pervasive Computing and Interaction (TPCI)* 1, 1 (2019), 3–9.

[108] Sangki Yun, Yi-Chao Chen, Huihuang Zheng, Lili Qiu, and Wenguang Mao. 2017. Strata: Fine-grained acoustic-based device-free tracking. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys'17)*. ACM, 15–28.

[109] Yunze Zeng, Parth H. Pathak, and Prasant Mohapatra. 2016. WiWho: WiFi-based person identification in smart spaces. In *Proceedings of the 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN'16)*. IEEE, 1–12.

[110] Youwei Zeng, Dan Wu, Jie Xiong, Enze Yi, Ruiyang Gao, and Daqing Zhang. 2019. FarSense: Pushing the range limit of WiFi-based respiration sensing with CSI ratio of two antennas. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 3, 3 (2019), 1–26.

[111] Chi Zhang, Josh Tabor, Jialiang Zhang, and Xinyu Zhang. 2015. Extending mobile interaction through near-field visible light sensing. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (MobiCom'15)*. ACM, 345–357.

[112] Cheng Zhang, Fan Yang, Gang Li, Qiang Zhai, Yi Jiang, and Dong Xuan. 2018. MV-Sports: A motion and vision sensor integration-based sports analysis system. In *Proceedings of the 37th Annual IEEE International Conference on Computer Communications (INFOCOM'18)*. IEEE, 1070–1078.

[113] Feng Zhang, Chenshu Wu, Beibei Wang, Hung-Quoc Lai, Yi Han, and K. J. Ray Liu. 2019. WiDetect: Robust motion detection with a statistical electromagnetic model. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 3, 3 (2019), 1–24.

[114] Lei Zhang, Meiguang Liu, Liangfu Lu, and Liangyi Gong. 2018. Wi-Run: Multi-runner step estimation using commodity Wi-Fi. In *Proceedings of the 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON'18)*. IEEE, 1–9.

[115] Lei Zhang, Zhirui Wang, and Liu Yang. 2019. Commercial Wi-Fi based fall detection with environment influence mitigation. In *Proceedings of the 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON'19)*. IEEE, 1–9.

[116] Shigeng Zhang, Chengwei Yang, Xiaoyan Kui, Jianxin Wang, Xuan Liu, and Song Guo. 2019. ReActor: Real-time and accurate contactless gesture recognition with RFID. In *Proceedings of the 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON'19)*. IEEE, 1–9.

[117] Yang Zhang, Chouchang (Jack) Yang, Scott E. Hudson, Chris Harrison, and Alanson Sample. 2018. Wall++: Room-scale interactive and context-aware sensing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI'18)*. ACM, 1–15.

[118] Mingmin Zhao, Fadel Adib, and Dina Katabi. 2016. Emotion recognition using wireless signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking (MobiCom'16)*. ACM, 95–108.

[119] Mingmin Zhao, Tianhong Li, Mohammad Abu Alsheikh, Yonglong Tian, Hang Zhao, Antonio Torralba, and Dina Katabi. 2018. Through-wall human pose estimation using radio signals. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18)*. IEEE, 7356–7365.

[120] Mingmin Zhao, Yonglong Tian, Hang Zhao, Mohammad Abu Alsheikh, Tianhong Li, Rumen Hristov, Zachary Kabelac, Dina Katabi, and Antonio Torralba. 2018. RF-based 3D skeletons. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication (SIGCOMM'18)*. ACM, 267–281.

[121] Yue Zheng, Yi Zhang, Kun Qian, Guidong Zhang, Yunhao Liu, Chenshu Wu, and Zheng Yang. 2019. Zero-effort cross-domain gesture recognition with Wi-Fi. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys'19)*. ACM, 313–325.

[122] Tong Zhu, Qiang Ma, Shanfeng Zhang, and Yunhao Liu. 2014. Context-free attacks using keyboard acoustic emanations. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS'14)*. ACM, 453–464.

[123] Han Zou, Yuxun Zhou, Jianfei Yang, Weixi Gu, Lihua Xie, and Costas J. Spanos. 2018. WiFi-based human identification via convex tensor shapelet learning. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI'18)*. AAAI, 1711–1718.

[124] Yongpan Zou, Jiang Xiao, Jinsong Han, Kaishun Wu, Yun Li, and Lionel M. Ni. 2017. GRfid: A device-free RFID-based gesture recognition system. *IEEE Transactions on Mobile Computing (TMC)* 16, 2 (2017), 381–393.

Received 15 February 2021; revised 4 February 2022; accepted 14 March 2022